

Improving Search Relevance Feedback through Human Centered Design

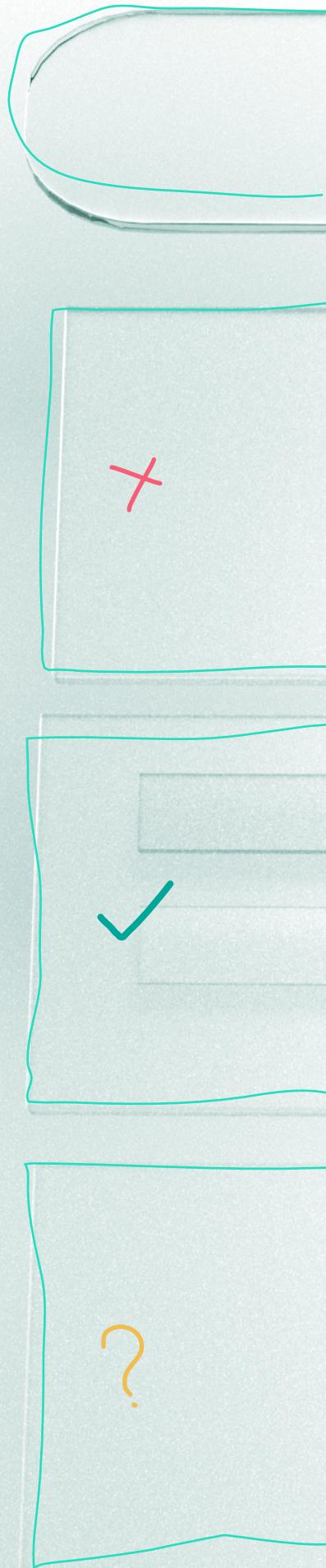
DfI Master Thesis

Shengfeng Gu

August 2020



mytomorrows





***“The more you know,
The more you don’t know.”***

- Aristotle



The Design for Interaction (DfI) is a 2-year Master's programme at the faculty of Industrial Design Engineering (IDE), it focuses on the ways in which people and products interact: how does a user understand, use, and experience a product? This raises the question of how designers can conceptualize products that are relevant to the user.

Chair

Alessandro Bozzon (DE)
A.Bozzon@tudelft.nl

Mentor

Derek Lomas (ID)
J.D.Lomas@tudelft.nl

myTomorrows

myTomorrows is an Amsterdam-based pharma-tech company, it uses AI and supporting services to improve the access to and recruitment of novel treatment in development (e.g. CT, EAP). By engaging more potential participants and matching them to clinical trials, myTomorrows can provide access to life-saving new medicines for patients and advance the pace of medical discovery. <https://mytomorrows.com/>

Company mentor

Zoltán Szlávik (AI lead)
zoltan.szlavik@mytomorrows.com

2nd Company mentor

Robert-Jan Sips (CTO)
robert-jan.sips@mytomorrows.com

Author

Shengfeng Gu | Msc. Design for Interaction
Email: Goooleon@gmail.com
Website: www.shengfenggu.com

Preface

This project is the final project of my two-year study in the Master track Design for Interaction at Delft University of technology. It marks the end of my Master's study and is possibly the last milestone of my student life (If not pursuing a Doctorate or another Masters's). The six months of work collaborating with myTomorrows was pleasant, enabling me to think more critically. More importantly, I am honored to work with those talented people at myTomorrows and be a part of the team striving to mitigate the gap our current medical systems fail to fill, especially during the time of the COVID-19 pandemic.

Firstly, I would like to appreciate my supervisory team for the patience and generous guidance. Thank you all for this enjoyable and unforgettable experience! Thank you for the trust and endless support that encouraged me to freely explore the possibilities in this project.

Thank you, Alessandro, for introducing me to the graduation opportunity and challenging me with your critical thoughts. It helped me form a more in-depth understanding of the project and enlightened directions.

Thanks to Derek for often guiding me through the project with specific and actionable steps. Besides, I appreciate it that you always share a bunch of inspirational materials through email and WhatsApp messages.

Zoltán, the company mentor, thank you for sharing the insights of your expertise and the patience of explaining all technical details that are not directly relevant to this project but inspiring.

A special thanks to all people who participated in the interviews, creative sessions, and evaluation of this project. I could never have achieved this project without the generous help from my fellow design students, myTomorrows colleagues, and healthcare professionals from Erasmus MC, China, and Brazil. Thank you RJ, Melissa, Yasmine, and Perry for giving early-stage input and feedback during concept development.

Last but not least, I would express my sincere gratitude to my friends and family for caring about me and supporting me throughout the project from afar. Support from you is the strongest motivation for me to keep trying and achieve better.

Please enjoy the read, and I hope this work will be inspiring!

Shengfeng Gu.

Table of contents

10 Executive summary

12 project overview

13 Glossary and abbreviation

Chapter 01 - Introduction

16 1.1 Unmet medical needs and investigational drugs

16 1.1.1 Patient with unmet medical needs

17 1.1.2 Drug development

19 1.2 myTomorrows Introduction

19 1.2.1 Mission

21 1.2.2 AI-powered treatment search engine

21 1.2.3 Vision

22 1.3 Project challenge

22 1.3.1 An intersection of HCD, Health & Wellbeing, and AI

22 1.3.2 Problem definition

23 1.3.3 Research questions

23 1.3.4 Approach

Chapter 02 - AI & Search, theoretical background

26 2.1 Search foundations

26 2.1.1 Search and information retrieval

26 2.1.2 Web search engine

27 2.1.3 Search and AI

28 2.2 The design of search

28 2.2.1 The anatomy of search

28 2.2.3 Search design guideline

30 2.3 Search as information seeking

30 2.3.1 A cognitive model

31 2.3.2 Information seeking process

32 2.3.3 Information seeking stages and emotion

33 2.3.4 Seeking behaviors

34 2.3.5 Information foraging as a search strategy

36 2.4 Relevance feedback foundations

36 2.4.1 Relevance and dimension

37 2.4.2 Feedback

37 2.4.3 Relevance feedback

38 2.4.4 Implicit feedback vs Explicit feedback

39 2.5 Theories on engagement

39 2.5.1 Engagement theory

40 2.5.2 Nudging theory

40 2.5.3 Persuasive theory and behavioral change

41 2.5.4 Persuasion cues

41 2.5.5 Theories in motivation

44 2.6 Perceived trust in search

44 2.6.1 What is trust?

44 2.6.2 Trusting dimensions and models

45 2.6.3 System trust in search

46 Key takeaways chapter 02

Chapter 03 - The current, exploratory research

50 3.1 Product analysis

50 3.1.1 Search decomposition

52 3.1.2 Heuristic evaluation

54 3.2 Future vision

54 3.2.1 future vision interviews

55 3.2.2 vision clusters

57 3.2.3 Clinicaltrial.gov Modernization

58 3.3 User research (healthcare professionals)

58 3.3.1 Empathizing healthcare professionals

59 3.3.2 HCP journey to CT/EAP

60 3.3.3 CT/EAP in HCPs' eyes

61 3.3.4 HCPs expectations on myT treatment search

62 3.3.5 HCP search personas

63 3.4 Building blocks of feedback

63 3.4.1 Treatment search flow

64 3.4.2 Reasons behind relevance feedback

66 3.4.3 How HCPs perceive feedback provision

66 3.5 Feedback collection in practice

66 3.5.1 Feedback purpose

67 3.5.2 Triggering position

70 Key takeaways chapter 03

Chapter 04 - Insights synthesis, envisioning the future

74 4.1 Design goal

74 4.1.1 Design goal breakdown

76 4.2 Design criteria and requirements

Chapter 05 - Future search conceptualization

82 5.1 Creative session for design ideation

82 5.1.1 Approach

83 5.1.2 Creative session results

86 5.2 Baseline search construction

86 5.2.1 Information architecture reconstruction

86 5.2.2 A visualized journey

88 5.2.3 Five user flows

90 5.3 Feedback ideas generation

90 5.3.1 Feedback collection mechanism

90 5.3.2 Nine feedback ideas

95 5.4 Usability testing and ideas speed dating

95 5.4.1 Methods

99 5.4.2 Study results

101 5.4.3 Insights from the speed dates

103 5.5 Feedback enhanced search conceptualization

103 5.5.1 Brainstorming session on feedback collection

104 5.5.2 Feedback ideas by stage

106 5.5.3 Iterated baseline search

114 5.5.4 Three synthesized feedback concepts

120 Key takeaways chapter 05

Chapter 06 - Concept evaluation and generalization

124 6.1 Concept evaluation with target users

124 6.1.1 Methods

126 6.1.2 Test results

132 6.1.3 Insights on relevance feedback

129 6.2 A guide for designing explicit feedback collection

135 6.2.1 A four-stage design guide

141 Key takeaways chapter 06

Chapter 07 - Discussion and recommendations

144 7.1 Addressing the research question

144 7.1.1 research question

145 7.2 Project contributions

145 7.2.1 Contribution to myTomorrows practice and Health & Wellbeing

146 7.2.2 Contribution to Design practice

146 7.2.3 Contribution to AI discipline

146 7.3 Limitations and recommendations

146 7.3.1 Limitations of this research

147 7.3.2 Recommendations for development at myTomorrows

147 7.3.3 Recommendations for future research

149 7.4 personal reflection

150 Bibliography

152 Appendices

Executive summary



Artificial intelligence (AI) is expected to play a transformational role in health and wellbeing. Search (i.e. information retrieval) technologies already play a significant role in healthcare research and practice (e.g. supporting continuing medical education and systematic reviews, Byron et al., 2012). Relevance feedback in Search is vital for system evaluation and improvements. However, in small user scale contexts, the exploitation of user behaviors may not infer valid relevance judgments. Therefore, engaging users to provide such feedback explicitly is essential.

In collaboration with myTomorrows, an Amsterdam-based pharma-tech company, this Master thesis aims to improve the user experience of an AI-powered treatment Search by engaging healthcare professionals (HCPs) in providing relevance feedback on search results (e.g. Clinical Trials and Expanded Access Programs). The outcome comprises a conceptual (user interface) Search enhanced with three explicit relevance feedback collection concepts, and a generalized guide for designing explicit feedback collection in text-base Search.

The research starts with an introduction to the general background of unmet medical needs and the project's challenges, followed by the research path towards the two outcomes. With the research question 'How to engage healthcare professionals to provide trustworthy relevance feedback on search results in myTomorrows Search', this project employs Human Centered Design methods to seek answers consisting of theoretical and exploratory research.

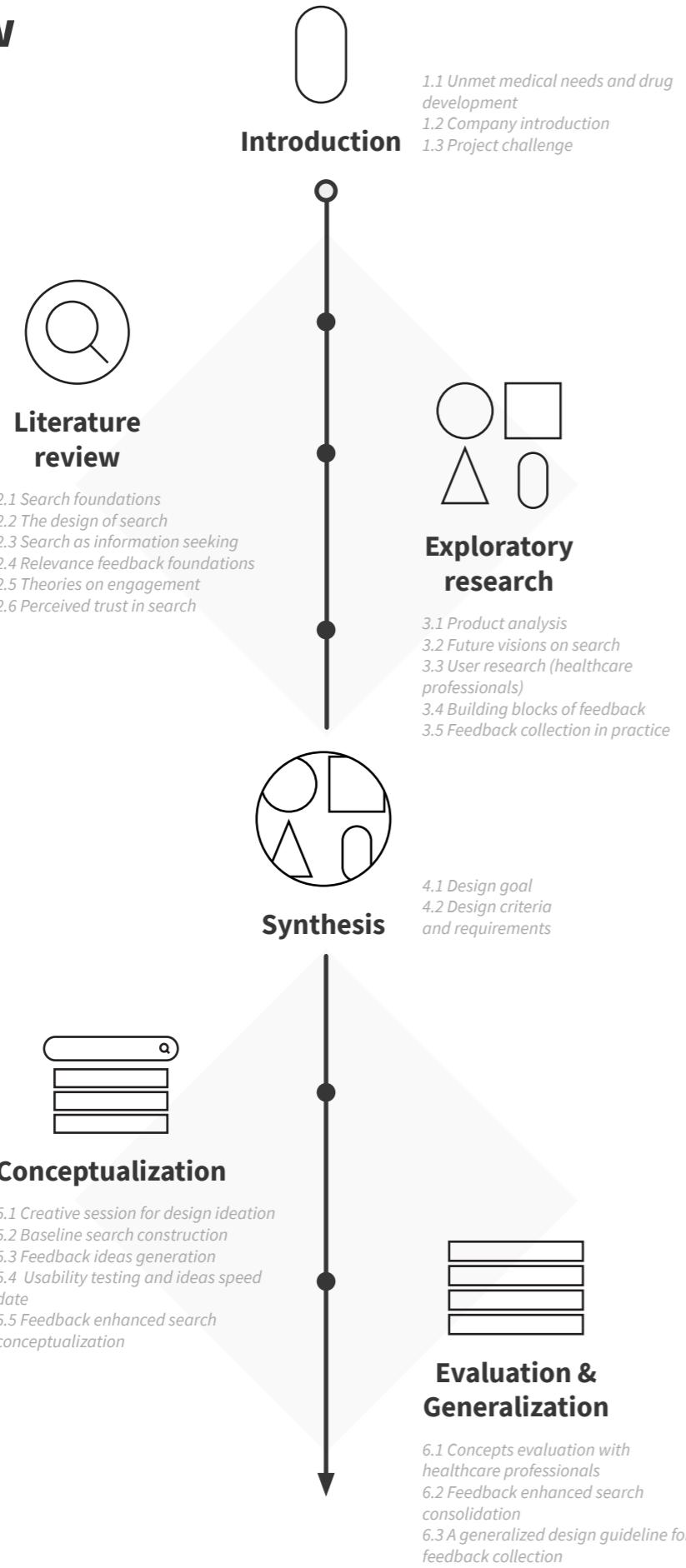
The theoretical study (chapter 2) investigates the literature on search systems, information seeking processes, and relevance feedback. However, the finding that users are generally reluctant to provide feedback explicitly expands the research scope to the realm of motivational theories and persuasive behavioral models. On top of the theoretical foundations, the exploratory study (chapter 3) dives into the current user experience of myTomorrows Search and identifies the future vision through the lens of HCPs and myTomorrows by interviewing, observing, and role-playing. Moreover, it discovers the determining factors of

relevance judgment and wherein a search process such judgments could be made (SERP and result content page). Generally, there are three types of relevance judgment: relevant, irrelevant, and uncertain. For HCPs to make irrelevant ones, it could be any reason belonging to system failure, insufficient expression of information needs, or flaw in source data. The lack of knowledge or inadequate information is the primary reason for uncertain judgment. Additionally, feedback collection designs in practice were looked into for inspiring ideation.

With the research insights, a design goal is formulated (chapter 4) with four criteria (Trust, Guidance, Product vision, and Contribution) and specific design requirements. They guided the design (chapter 5) of the baseline Search concept (Trust, Guidance, and Product vision), on top of which nine feedback collection ideas were built and speed dated with peers and myTomorrows employees as a primary source of insights. The learnings led to three relevance feedback collection concepts (Pre-screener, Reminder, Hearty) embedded with different motivators (personal utility, altruism, and enjoyment), which were evaluated (chapter 6) with nine HCPs from three countries (Netherlands, China, Brazil). The results indicate that the first concept (Pre-screener) slightly outperforms the other two in terms of rating, preference, and heuristics, but the feedback data collected could be short-sighted. In comparison, the second concept (Reminder) is generally more accepted and is perceived as more motivating because it actively nudges users. However, the third concept (Hearty) is considered inappropriate for the context despite the fact that two HCPs showed interest in it.

To sum up, for future development (Chapter 7) of explicit relevance feedback collection in myTomorrows Search, it is recommended to combine the essentials of all three concepts. For other cases of engaging users to provide feedback, the generalized design guide could shed light on research and design with a four-stage process and ten recommendations. However, the guide stays a generalization of this specific myTomorrows Search case and needs further validation.

project overview



Glossary and abbreviation

myT	myTomorrows
HCP	Healthcare professional
CT	Clinical trials
EAP	Expanded Access Program
SERP	Search engine result page
HCD	Human-centered design
UX	User Experience
AI	Artificial Intelligence
ML	Machine Learning
IR	Information Retrieval

Search (equivalent to information retrieval) is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources.

Relevance feedback user feedback of relevance judgment on the information (i.e. CT/ EAP documents) retrieved by the search system to a certain information need.

Human Centered Design an approach to problem solving, commonly used in design and management frameworks that develops solutions to problems by involving the human perspective in all steps of the problem-solving process.

Internal HCP myTomorrows employee with medical education background and is responsible for dealing with medical related work (e.g. generating treatment search reports)

External HCP any member (not working for myTomorrows) of the medical (e.g. physician), pharmacy or nursing professions or any other person who in the course of his or her professional activities may prescribe, administer or dispense to an end-user a medicinal product.

Speed dating a design method for rapidly exploring application concepts and their interactions and contextual dimensions without requiring any technology implementation.

01

Introduction



This chapter covers a comprehensive introduction to the project. Starting from a broad perspective of the unmet medical needs context, it lays the foundation for what myTomorrows aims to resolve and achieve. Followed is the company introduction, outlining a holistic overview of myTomorrows business and services for a deeper understanding of where this thesis positions and contributes. The last section extensively describes the assignment and the challenges this project faces: the urge for 1) overall UX improvement and 2) engaging users to provide explicit relevance feedback in Search.

Chapter overview

- 1.1 Unmet medical needs and drug development
- 1.2 Company introduction
- 1.3 Project challenge



1.1 Unmet medical needs and investigational drugs

The screenshot shows the Cancer Research UK website. At the top, there is a logo and the slogan "Together we will beat cancer". Below the header, there are navigation links for "About cancer", "Get involved", "Our research", "Funding for researchers", "Shop", and "About us". A search bar and a "Donate" button are also present. The main content area shows a forum post titled "I'm running out of treatment options" by a user named "Hope333". The post includes a profile picture, the date "28 Mar 2019 22:48", and 8 replies. The post content discusses the user's cancer journey and current treatment challenges.

Figure 1. Screenshot of a help post by a patient with breast cancer (source: Cancer Research UK) <https://www.cancerresearchuk.org/about-cancer/cancer-chat/thread/im-running-out-of-treatment-options>

Unmet medical needs mean a condition for which there exists no satisfactory method of diagnosis, prevention or treatment in the Union or, even if such a method exists, in relation to which the medicinal product concerned will be of major therapeutic advantage to those affected.

(European Medicines Agency, 2006)

1.1.1 Patient with unmet medical needs

There are approximately 70 million patients in the world with an unmet medical need, such as patients with rare diseases (While more than 7,000 rare diseases have been identified, only 5 percent have treatments.) or those who have run out of standard treatment options (Figure 1).

According to the research conducted in the Netherlands (Bunnik, E. M., & Aarts, N. 2019), patients with unmet medical needs may seek non-standard treatment options (Bunnik, E. M., & Aarts, N. 2019) such as the use of investigational drugs in a clinical trial (CT) setting or expanded access programs (EAPs).

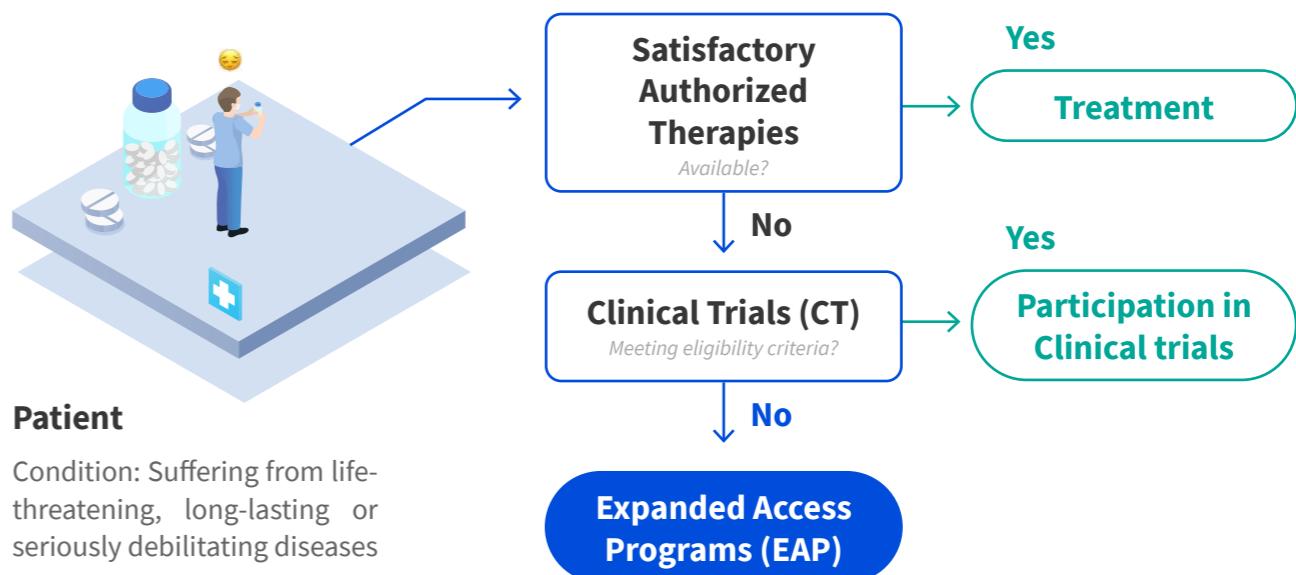


Figure 2. Pathway to compassionate use program. (image sources: Balasubramanian et al., 2016, adapted by author)

A pathway to CT/EAP

Figure 2 illustrates the pathway to access investigational medicines through compassionate use programs (i.e. Expanded Access programs) for patients suffering from severe or enervating diseases (Balasubramanian et al., 2016). Clinical trials are protocol-driven where patients have to meet specific eligibility criteria, while EAP allows patients without considering any requirements, but only when there are no CTs of such treatments or he/she is not eligible to be included in CTs.

Obstacles to access CT/EAP

However, patients barely have medical knowledge or experience with investigational drugs, although some were searching for non-standard treatment options. Patients have high expectations for their treating physicians and assume them to be aware of and inform patients about non-standard treatment options. Also, patients may prefer their treating physicians discuss such treatment options with them, regardless of the medical knowledge barriers. Besides, patients are careful about the risks and benefits, including safety, efficacy, side effects, drug-drug interaction, and the maintaining of good quality in life. The other major obstacle for patients with unmet medical needs is the inaccessibility of new treatment (Mehta, A. 2008).

1.1.2 Drug development

Drug development is strictly regulated and has to go through a costly and time-consuming process of research and approval before it can be registered as a commercial product or used as a standard treatment. According to FDA regulation, the drug development process (Figure 3) comprises five steps in which clinical research requires 4 phases of clinical trials with human participants.

Challenges in drug development

Carrying out clinical research is demanding both for the length of study and the number of study participants, which influences the translational process from research into standards of care (Penberthy, L. T. et al. 2012). However, the capacity of recruiting participants (Haidich et al. 2001) determines the success of CT in a given time frame. It is a significant challenge in clinical research (Pressler, T. R et al., 2012).

A previous study (DiMasi, J. A. et al. 2003) has shown that a 25% reduction in phase lengths could lower the capitalized total cost per approved drug by 16%. Ultimately, the reduced cost and increased productivity could result in more innovation in drug development and new treatments reaching patients earlier (DiMasi, J. A. 2002), saving or prolonging lives.

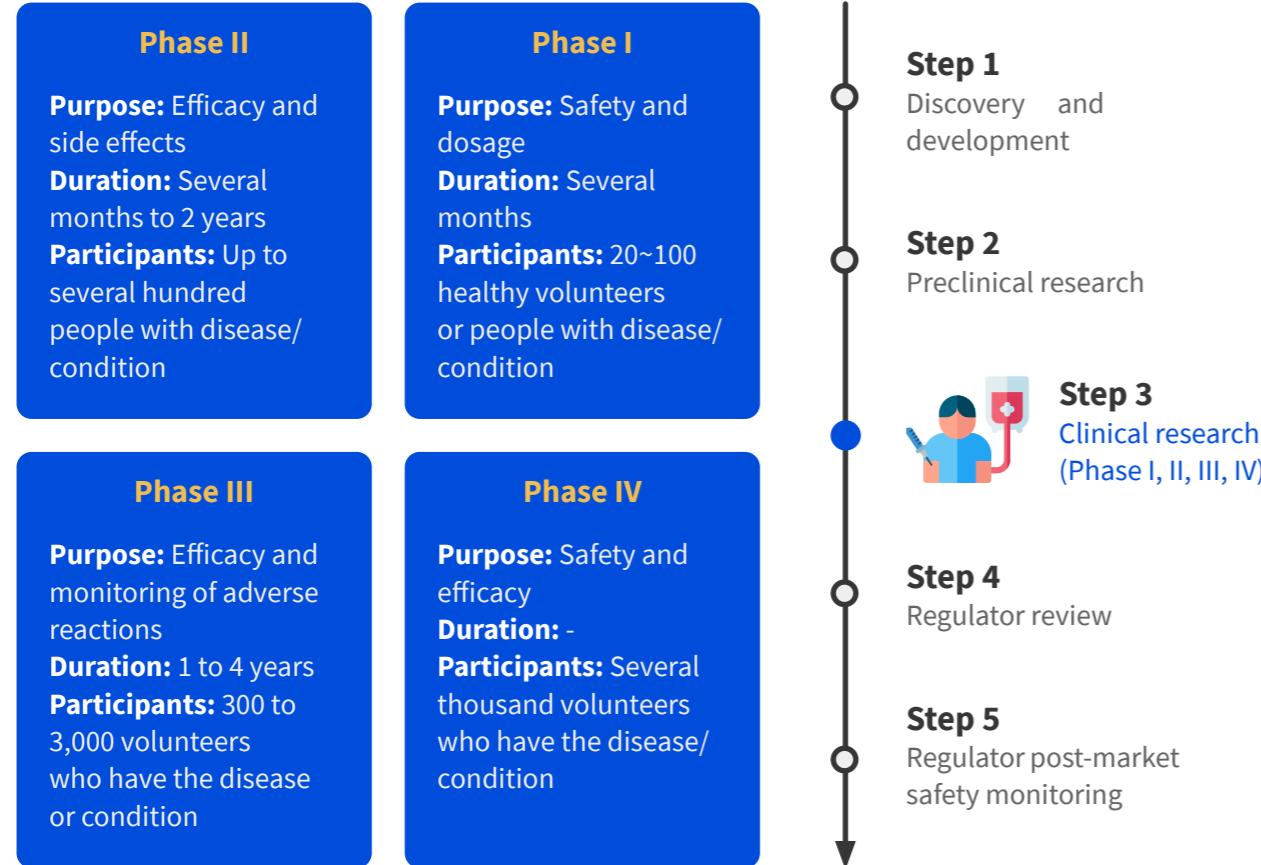


Figure 3. Drug development process (Image source: U.S FDA, adapted by author)

Clinical Trial (CT)

CT (Figure 3) is a type of clinical research in which human participants are assigned to groups that receive one or more intervention/treatment (or no intervention) to evaluate the effects of the interventions on biomedical or health-related outcomes. As the purpose of CT is research rather than treatment, inclusion and exclusion criteria are rigorous that not all patients are eligible to participate.

Expanded Access Program (EAP)

Expanded Access Program (EAP), also called compassionate use, is a way for patients with serious diseases or conditions who cannot participate in a clinical trial to gain access to a medical product that has not been approved by the regulator. EAPs are rarely occurring (100-200 requests in the Netherlands in 2014) but see a significant rise over the past few years (Bunnik,

E. M. et al. 2018), 92% increase of compassionate access as reported by FDA in the United States.

At the moment, access to investigational products is available only to the lucky few who know about it, can draw attention to their plight, navigate the process of requesting it, and who are given access by drug companies, either by the company's own decision or as a result of pressure brought by an advocacy campaign or powerful person

(Caplan, A. L., & Bateman-House, A. 2015).

1.2 myTomorrows Introduction

myTomorrows (myT) is an Amsterdam-based pharma-tech company, and it uses Artificial intelligence (AI) and supporting services (Figure 4, 5) to improve access to and recruitment of treatment in development (e.g. CT, EAP). By engaging more potential participants and matching them to Clinical Trials or EAPs, myTomorrows can provide access to life-saving new medicines for patients and advance the pace of medical discovery.

1.2.1 Mission

Ten years or more, that is how long it can take for a medicine to be approved for the market. Doctors and their patients often do not have access to a medicine during the approval period, even though it may already show promising results. **myTomorrows is committed to improving access to treatment in development.** (myTomorrows, 2020)

"We want to ensure that patients don't miss out on treatment options because of a lack of information and understanding or due to administrative barriers."



Figure 4. myTomorrows service funnel (image source: myTomorrows)



Finding out if any treatment options may be available can be difficult.

All around the world, new medicines are constantly being developed. However, these medicines might not yet be available and are often difficult to access. Our medical team identifies emerging treatments using the latest technology. Here is how it works.

Figure 5. Screenshot of myTomorrows website homepage (image source: myTomorrows) <https://mytomorrows.com/en/>

myTomorrows internal search																	
Patient																	
Search for term in result...																	
Patient ID																	
Gender																	
All	Male	Female															
Age																	
Condition																	
Select a condition																	
Coronavirus																	
Condition CUI C0206750																	
select all																	
Coronavirus Infections (158)																	
Severe Acute Respiratory Syndrome (80)	<input checked="" type="checkbox"/>																
Middle East Respiratory Syndrome (7)	<input checked="" type="checkbox"/>																
Pneumonia caused by Human coronavirus (disorder) (2)	<input checked="" type="checkbox"/>																
flip (?)	<input checked="" type="checkbox"/>																
Location																	
Worldwide OR:																	
Specific location																	
NCT0469609	Clinical Trial	Mucous Fistula Refeeding Reduces the Time From Enterostomy Closure to Full Enteral Feeds (MUCous Fistula REfeeding ("MUC-FIRE"))	Status	Recruiting	18-06-2018	01-04-2022	<input type="checkbox"/>										
NCT00173459	Clinical Trial	Dynamic Profiles of Cytokine/Cytokine in Severe Acute Respiratory Syndrome	Status	Completed			<input type="checkbox"/>										
NCT00172263	Clinical Trial	The Interaction Between Severe Acute Respiratory Distress Syndrome Viral Proteins and Monocytes	Status	Withdrawn			<input type="checkbox"/>										
NCT00578825	Clinical Trial	A Multi-centre, Double-blinded, Randomized, Placebo-controlled Trial on the Efficacy and Safety of Lopinavir / Ritonavir Plus Ribavirin in the Treatment of Severe Acute Respiratory Syndrome	Status	Unknown status			<input type="checkbox"/>										
NCT03301090	Clinical Trial	A Safety, Tolerability, Pharmacokinetics and Immunogenicity Trial of Co-administered MERS-CoV Antibodies REGN3048 and REGN3051	Status	Completed	12-02-2018	19-01-2019	<input type="checkbox"/>										
NCT02845843	Clinical Trial	MERS-CoV Infection treated With A Combination of Lopinavir /Ritonavir and Interferon Beta-1b	Status	Recruiting	01-07-2016	01-12-2020	<input type="checkbox"/>										
NCT03225807	Clinical Trial	Implementation of Lung Protective Ventilation in Patients With Acute Respiratory Failure	Status	Recruiting	01-03-2016	01-12-2020	<input type="checkbox"/>										
NCT03615911	Clinical Trial	Safety, Tolerability and Immunogenicity of Vaccine Candidate MVA-MERS-S	Status	Completed	28-11-2017	10-05-2019	<input type="checkbox"/>										
NCT04119440	Clinical Trial	Randomized, Double-blind, Placebo-controlled, Phase Ib Study to Assess the Safety and Immunogenicity of MVA-MERS-S-DF-1	Status	Not yet recruiting	01-05-2020	01-12-2021	<input type="checkbox"/>										
NCT03399578	Clinical Trial	Safety and Immunogenicity of a Candidate MERS-CoV Vaccine (MERS001)	Status	Recruiting	14-03-2018	01-07-2021	<input type="checkbox"/>										
NCT04130594	Clinical Trial	Study of Safety and Immunogenicity of BVRS-GamVac	Status	Recruiting	07-11-2019	31-12-2020	<input type="checkbox"/>										
NCT04128059	Clinical Trial	Study of Safety and Immunogenicity of BVRS-GamVac-Combi	Status	Recruiting	06-11-2019	31-12-2020	<input type="checkbox"/>										
NCT01056185	Clinical Trial	Respiratory Virus Hospitalization Study (FLU 003 Plus)	Status	Recruiting	01-08-2009	01-12-2020	<input type="checkbox"/>										
NCT04170829	Clinical Trial	A Clinical Trial to Determine the Safety and Immunogenicity of Healthy Candidate MERS-CoV Vaccine (MERS002)	Status	Recruiting	17-12-2019	01-01-2021	<input type="checkbox"/>										
NCT04245531	Clinical	Development of a Simple, Fast and Portable Recombinase Aided Amplification Assay for 2019-nCoV	Status	Recruiting	01-01-	31-12-	<input type="checkbox"/>										

Figure 6. myTomorrows internal search (image source: myTomorrows)

Figure 7. myTomorrows external search - HCP portal. (image source: myTomorrows)
<https://search.mytomorrows.com/search/hcp>

1.2.3 AI-powered treatment Search

AI-powered Search plays an essential role (Figure 4, Inform) in bridging treatment information worldwide and is the key to scale myTomorrows business to fulfill its mission. myTomorrows has developed an internal Search and an external Search on the same technical infrastructure.

Internal search

The internal search (Figure 6) is developed for the internal medical team, who conduct searches and produce search reports based on requests. What differentiates the internal search is that it collects relevance feedback from internal HCPs, and it includes more CT registry databases.

Apart from the two differences mentioned above, the internal search has access to the patient document structuring system (PDS), which contains health record documents uploaded by patient users.

External search

The external search (Figure 9) incorporates two portals for patients and HCPs, aiming at empowering the target groups to reach and be aware of treatment options. There is a minor difference between the two portals in the user experience. For example, the difference in guidance information and the HCP portal has a saving (search result) feature.

1.2.4 Vision

Figure 88 shows the vision of myTomorrows service offering that it will play the role of a bridge connecting patients with unmet medical needs and pharmaceutical companies. There is the phased approach (Figure 99) aiming at a systemic change of offerings.

myTomorrows is a platform that opens up information access for all of our stakeholders, enabling earlier and better access to all possible treatment options.



Figure 8. myTomorrows service offering (image source: myTomorrows, adapted by author)

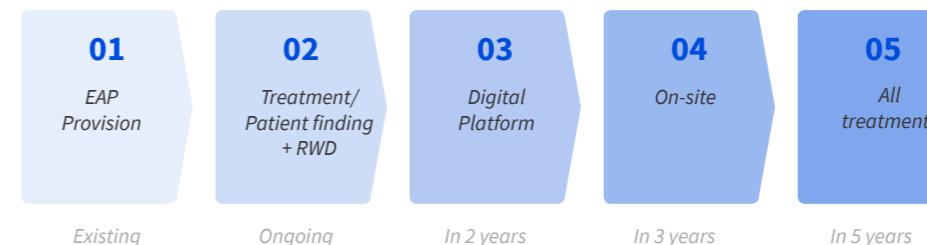


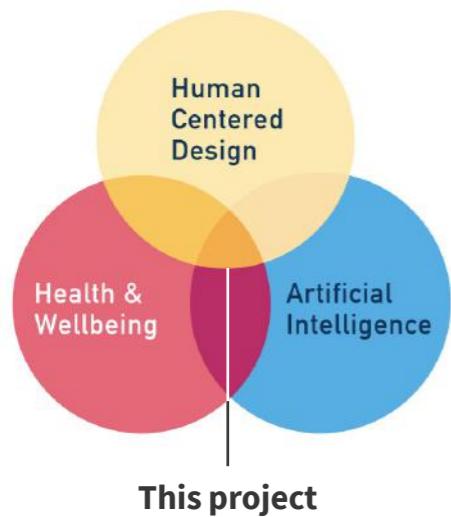
Figure 9. A phased approach towards systemic change (image source: myTomorrows, adapted by author)

1.3 Project challenge

1.3.1 An intersection of HCD, Health & Wellbeing, and AI

Artificial Intelligence (AI) is expected to play a transformative role in health and human wellbeing. Search and information retrieval technologies already play a significant role in healthcare research and practice (e.g., by supporting continuing education and systematic reviews, Byron et al., 2012).

How might Human-centered design methods support the design of AI systems within the field of health and wellbeing? This project will probe the problem myTomorrows search is exposed to through the lens of design and use of Human-centered design methods. By putting together the multidisciplinary knowledge and future visions, this project aims to improve the overall search experience and to unleash the full potential of myTomorrows AI recommendation system, thus, achieving a more clear and effective search experience.



1.3.2 Problem definition

Search, in the specific context of myTomorrows, is to match patients to relevant treatment(s) with

the support of healthcare professionals. Currently, however, two main problems are slowing down the process.

1) The unmet urge to improve the current search UX of guiding users throughout the system

First, the unmet urge to improve the current search UX of guiding users throughout the system. A search does not guarantee the retrieval of relevant information, and efforts are required even at the very beginning of forming the relevant query.

Search, as a labor-intensive process, manifests complex behavior patterns (Search patterns, 2010). It is intriguing to discover how myTomorrows Search could effectively and pleasantly guide users through the maze of information seeking to the relevant information.

2) The lack of trustworthy user judgment for system optimization

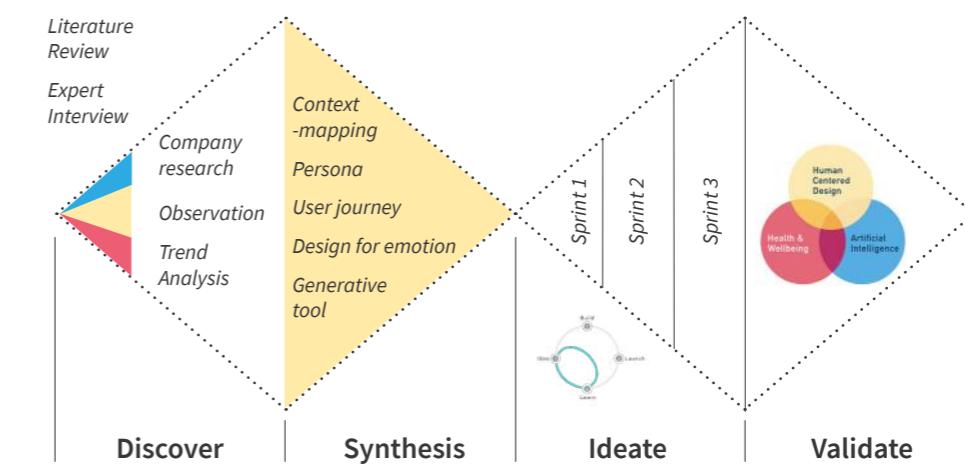
Another problem posed is the lack of trustworthy relevance judgment from end-users (Healthcare professionals). Implicit (Kelly et al., 2003) and explicit relevance feedback have been systematically studied and widely adopted, and recent research has gone creative in terms of interactive feedback collection (Juan, et al., 2017). At myTomorrows, however, it is not certain how to collect relevance data explicitly. Data sensitivity is the nature in the medical field, and it is difficult to draw the line where it meets the need for AI optimization and is also acceptable by users. Such issues lag the pace of AI advance and interrupt the workflows of the internal medical team and the AI research team.

Hence, the question is, how could relevance feedback with quality be collected and translated into useful training data in a data-sensitive and limited user scale context as myTomorrows?

1.3.3 Research questions

How to engage healthcare professionals to provide trustworthy relevance feedback on search results in myTomorrows Search?

- *Q1: What are the incentives for healthcare professionals to be engaged?*
- *Q2: What type of interaction approach would better trigger healthcare professionals to provide feedback?*
- *Q3: At what moment(s) should healthcare professionals be nudged or motivated to provide feedback?*
- *Q4: How to make sure the feedback provided by healthcare professionals is trustworthy?*



1.3.4 Approach

This project will run through four phases following the double diamond process. The first eight weeks will focus on collecting qualitative data, understanding the context, and building up search models for the next phase to take action.

The following two phases will emphasize on design sprints to deepen the understanding of guiding users and relevance feedback collection. In the end, a concrete concept will be designed and validated with end-users, together with a generalized framework or guideline of designing explicit relevance feedback collection in search.

02

Search & Feedback, theoretical background



To understand the foundations of search systems and how users could be engaged in providing relevance feedback for Search improvement, it is crucial to learn the theoretical fundamentals of search systems and user behaviors in Search. This literature review covers a broad spectrum, containing theories of information retrieval, Search design, information seeking process, and what constitutes relevance feedback. In addition to a technical perspective, the viewpoint focusing on end-users investigates the behavioral and psychological theories for understanding the determining factors in an engaging experience.

Chapter overview

- 2.1 Search foundations
- 2.2 The design of search
- 2.3 Search as information seeking
- 2.4 Relevance feedback foundations
- 2.5 Theories on engagement
- 2.6 Perceived trust in search



2.1 Search foundations

The search system (or information retrieval systems) is the media that users interact with, thus, this section looks into what makes up Information Retrieval systems and its latest developments of AI/ML integration.

2.1.1 Search and information retrieval

Search is a broad and ambiguous term, but in the modern parlance, it has tended to replace the meaning of “(information) retrieval” in information science. A web search engine such as Google, is a typical application of information retrieval systems.

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). (Manning, 2008)

In history, search has roughly been through 4 phases (Figure 10, Schatz, B. R. 1997) since 1960: Grand visions, Text search, Document search, and concept search. Each phase marks the technological advances in IR, and retrieving information becomes a more complicated process from the processing of query to the understanding of query semantic meaning.

Semantic Web and Semantic Search

Guha et al. (2003) interpreted the Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It connects information and builds relations between information to enable more

effective discovery, automation, integration, and reuse across various applications. In the Semantic Web, Knowledge graph is an increasingly critical component, and it serves as information hubs for general use as well as for domain-specific applications (McCusker, J. P. et al., 2018). For instance, Google deployed Knowledge graph 2012 to enhance search quality.

Semantic search is an application of the Semantic Web to search, and it attempts to augment and improve traditional search results (based on IR technology) by using data from the Semantic Web.

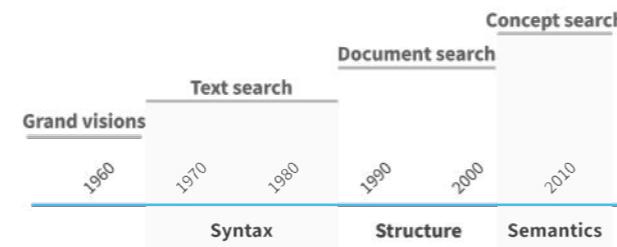


Figure 10. Rough timeline of the generations of information retrieval in digital libraries (image source: Schatz, 1997)

2.1.2 Web search engine

Search is a rather simple activity in terms of interaction by putting in a query and a simple click the search. Nevertheless, the mechanism of how search engines work is unknown to most users.

Figure 11 shows a composite picture of a web search engine, including the crawler and the indexer for web content and ads. The portion under the dashed line is internal to the search engine, and it is where AI technologies come into place for enhancing and improving search results.

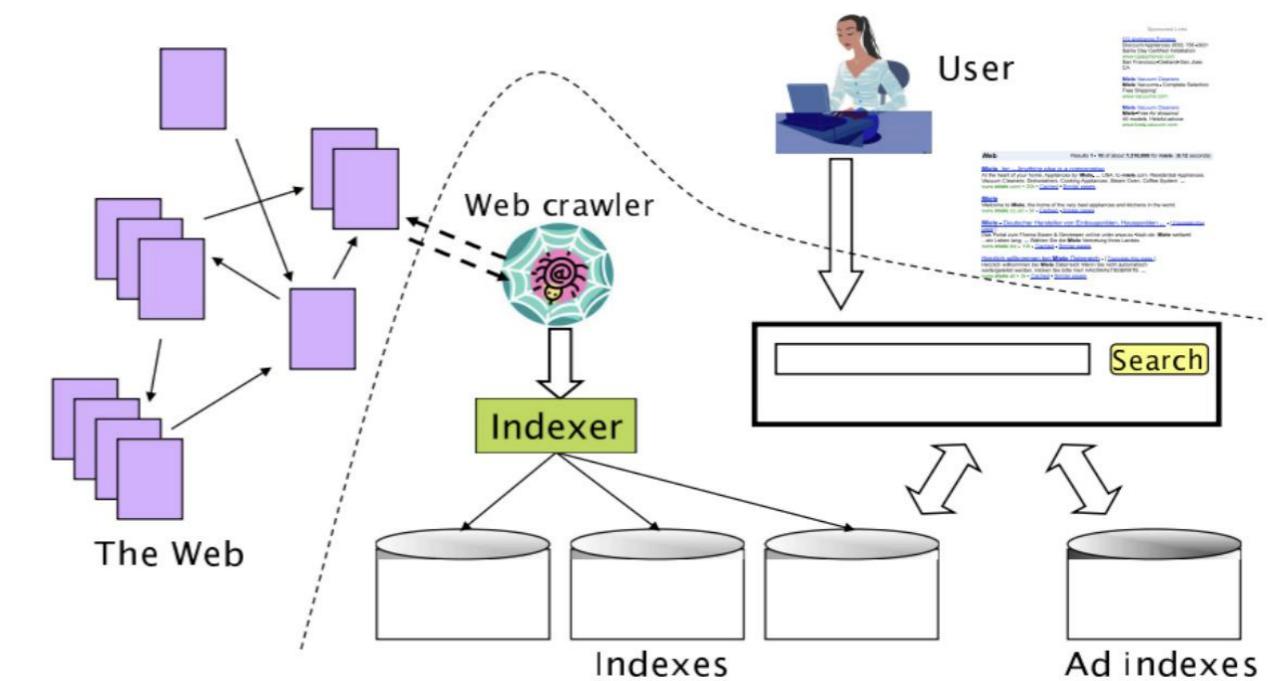


Figure 11. The various components of a web search engine (image source: Manning, 2009)

2.1.3 Search and AI

In the Artificial Intelligence (AI) domain, it is generally divided into Narrow AI (ANI) and the opposite General AI (AGI), some label it as Weak AI and Strong AI according to its capabilities.

General AI refers to a machine with the ability to apply intelligence to any problem, rather than just one specific problem, sometimes considered to require consciousness, sentience, and mind (Searle, J. R. 1980). In contrast, ANI describes AI systems that are specified to handle a singular or limited task. In this sense, a search engine powered by machine learning technologies (Figure 12) could be regarded as an application of AI. Typical usage of ML is recommender systems, and they could be found in many IR systems (e.g., Amazon).

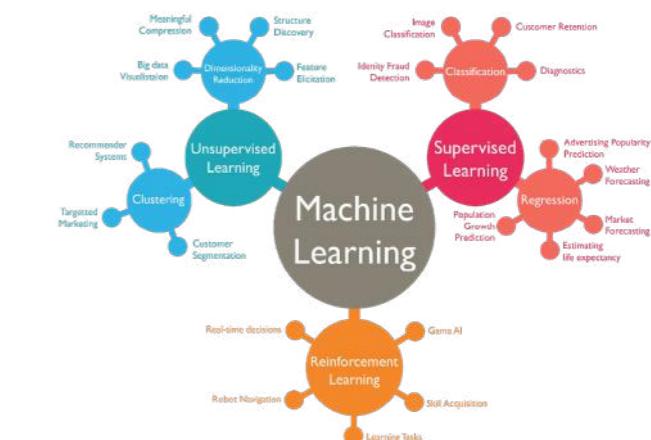


Figure 12. Machine learning taxonomy (image source: Oracle, 2018)

2.2 The design of search

A search system is commonly in the form of a graphical user interface (GUI), and this section aims to discover established standards or patterns in search design practices. Ultimately, to be aware of factors that should be considered while designing for a Search UI.

2.2.1 The anatomy of search

Morville (2010) maps out five elements (Figure 13) in search design: users, creators, content, engine, and interface. Each element features its factors that determine the design of search (engine). There are different types of users with different levels of expertise, knowledge, and expectations from a search system, such as recall versus precision. Most users view the engine as a black box, and they only query and look for results while interacting with the interface.

Design patterns in search

It has been recognized that there are ten common design patterns (Figure 14) in search systems:

Autocomplete (query suggestions)

Best first refers to the number and presentation of suggested links and their relationship to algorithmic results.

Federated Search involves the simultaneous search of multiple databases or collections. (e.g. In libraries, it lets users search multiple catalogs, collections, databases, and websites all at once.)

Faceted Search leverages metadata fields and values to provide users with visible options for clarifying and refining queries.

Advanced Search includes whatever simple search doesn't (e.g. Boolean)

Personalization

Pagination most queries produce too many results for one screen, pagination is a common solution.

Structured results helps users dig deeper into the data so users don't have to.

Actionable results

Unified Discovery Search rarely stands alone. In most contexts, users move between modes of searching, browsing, and asking.

2.2.3 Search design guideline

Based on eight desiderata for search user interfaces specified by Shneiderman (1997), Hearst (2009) synthesizes a search interface guideline that consists of seven aspects:

1. Offer Efficient and informative feedback

- Show search results immediately
- Show Informative Document surrogates and highlight query terms
- Allow sorting of results by various criteria
- Show query term suggestions
- Use relevance indicators sparingly
- Support rapid response

2. Balance user control with automated actions

- Rank ordering in web search
- Query transformations

3. Reduce short-term memory load

- Suggest the search action in the entry form
- Support simple history mechanisms
- Integrate navigation and search

4. Provide shortcuts

5. Reduce errors

- Avoid empty results sets
- Address the vocabulary problem

6. Recognize the importance of small details

7. Recognize the importance of aesthetics in design

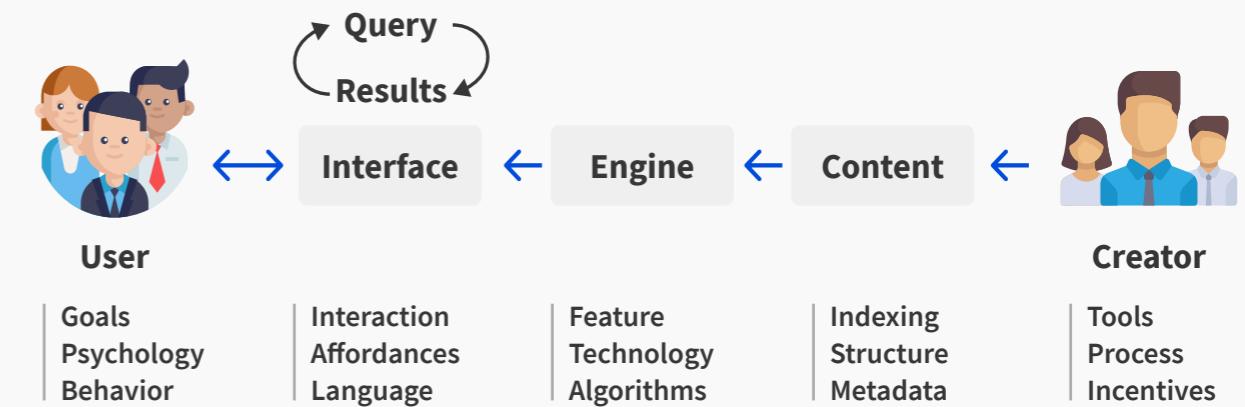


Figure 13. The anatomy of search (image source: Design patterns 2010)

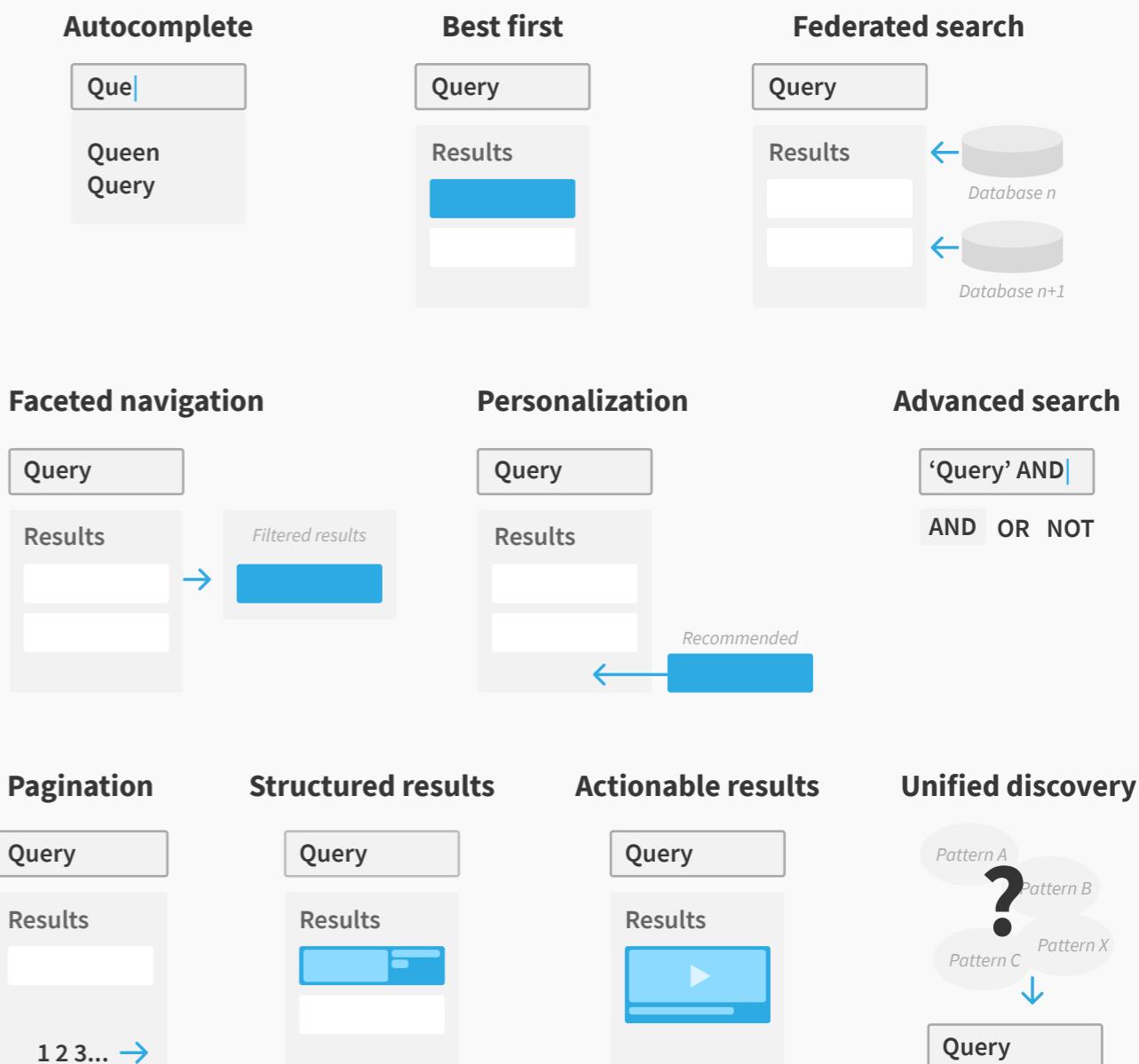


Figure 14. Search design patterns (image source: Search patterns 2010)

2.3 Search as information seeking

Innovations in technologies have brought human beings into an era of information explosion in multiple dimensions (Korth, H. F., & Silberschatz, A., 1997) such as breadth and amount of information, and the access to information on the web is much easier than ever before. However, seeking relevant and useful information is never an easy task to operate, instead, it is a complex process that consumes cognitive (Sutcliffe, A., & Ennis, M., 1998) and external resources.

2.3.1 A cognitive model

Seeking information is a high cognition consumption activity as Norman's model (Figure 15) of general task performance (1988). Psychological interpretation of such activity comprises the formulation of a goal to be achieved, task execution, and evaluation of execution gains. The gap between what was intended and what was achieved as the gulf of execution, and the challenge of determining whether or not one's goal have been met as the gulf of evaluation, meaning that such activity is dynamic and requires constant refinement and adjustment.

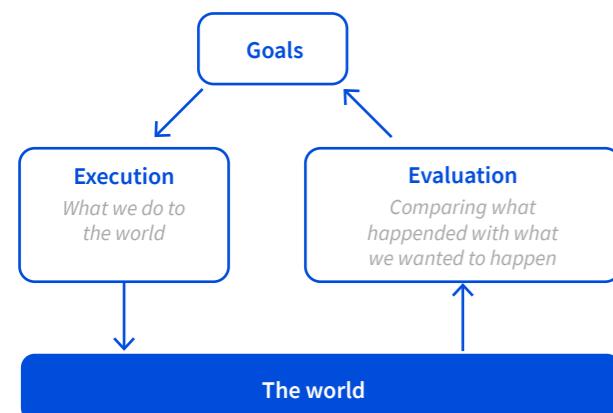


Figure 15. Cognitive execution-evaluation model (image source: Norman, adapted by author)

Search intentions

Broder (2002) came up with the taxonomy of web searches, where three classes were identified based on different intentions for searching:

Navigational: The immediate intent is to reach a particular site.

Informational: The intent is to acquire some

information assumed to be present on one or more web pages.

Transactional: The intent is to perform some web-mediated activity.

On top of the intentions, the researcher also identified corresponding search queries that may infer those intentions.

Navigational queries: The purpose of such queries is to reach a particular site that the user has in mind, either because they visited it in the past or because they assume that such a site exists.

Informational queries: The purpose of such queries is to find information assumed to be available on the web in a static form. No further interaction is predicted, except reading. By static form we mean that the target document is not created in response to the user query. This distinction is somewhat blurred since the blending of results characteristic to the third generation search (which attempts to blend data from multiple sources in order to try to answer 'the need behind the query) engines might lead to dynamic pages.

Transactional queries: The purpose of such queries is to reach a site where further interaction will happen. This interaction constitutes the transaction defining these queries. The main categories for such queries are shopping, finding various web-mediated services, downloading various type of file (images, songs, etc), accessing certain data-bases (e.g. Yellow Pages type data), finding servers (e.g. for gaming) etc.

2.3.2 Information seeking process

Information-seeking is a special case of problem solving. It includes recognizing and interpreting the information problem, establishing a plan of search, conducting the search, evaluating the results, and if necessary, iterating through the process again.

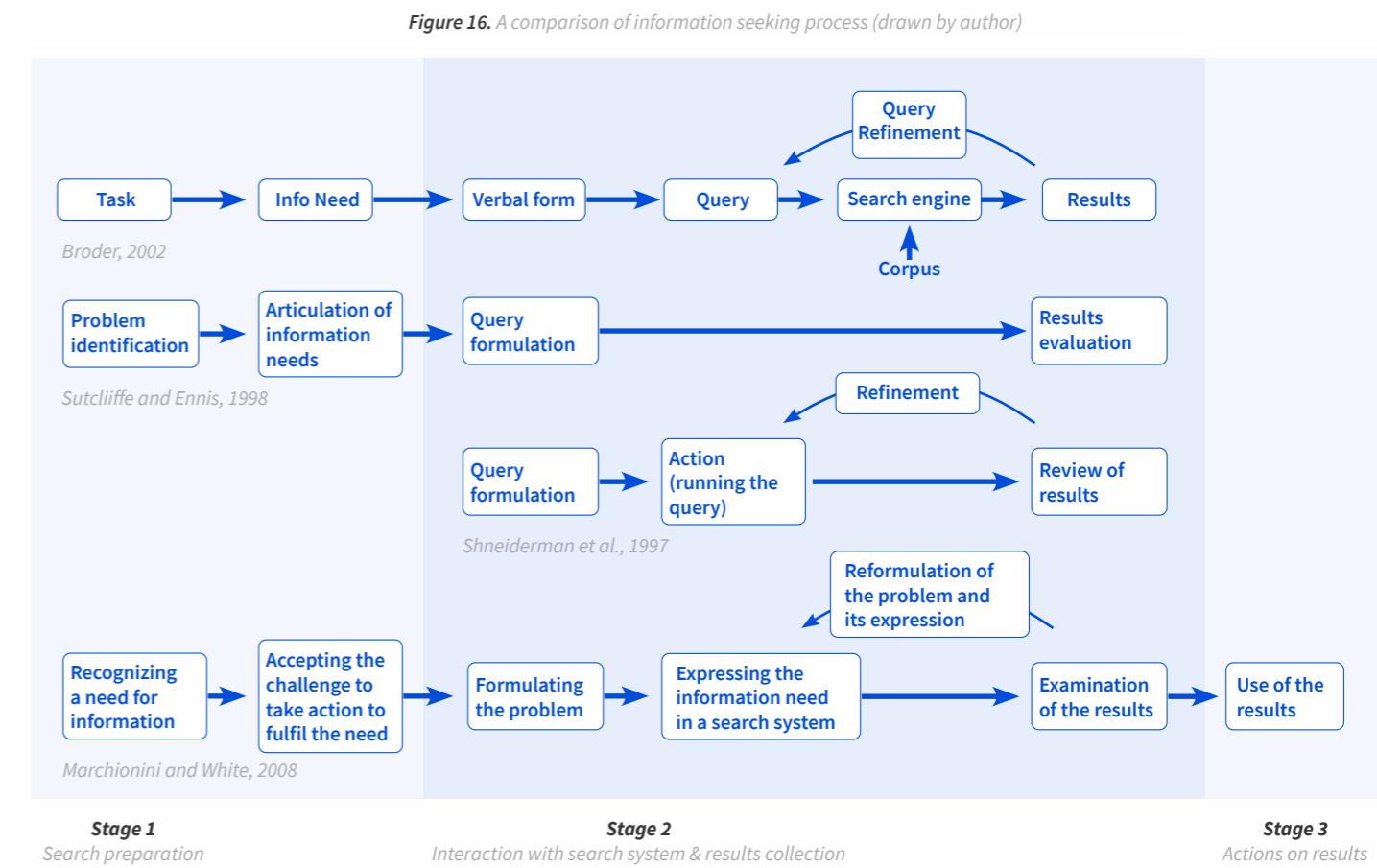
(Marchionini, G., 1989)

Researchers (Broder, 2002; Sutcliffe, A., & Ennis, M. 1998; Shneiderman et al., 1997; Marchionini, Gary, & White, R. 2007) have studied the information

seeking process and proposed different variations of information seeking models.

Figure 16 shows a comparison of information seeking models. They vary in the scope and steps, but they all share a common pattern that seeking information is a repetitive and iterative process where (query) refinement is an essential step. It is supported by other studies (Patterson et al., 2001; Jonker et al., 2005) that information quality and accuracy improve as searchers spend more time digging out more relevant documents containing information matched with a searcher's real information needs. This finding correlates to another theory of sensemaking (Russell et al., 1993).

Most models stop at the search results collection and examination stage, excluding the use of search results. Nevertheless, the similarities they share reveal three stages: Stage 1: Search preparation, Stage 2: Interaction with search system & Results collection, and Stage 3: Actions on results.



2.3.3 Information seeking stages and emotions

Kuhlthau (1991) researched information seeking from the searchers' perspective, and proposed a stage model of information seeking with the involvement of emotional changes (Figure 18). The model includes 6 stages of Initiation, Selection, Exploration, Formulation, Collection, and Presentation. Among all stages, Formulation marks the turning point in the search process that a focused perspective on the topic emerges. Searchers' emotion states fluctuate along the way of seeking information.

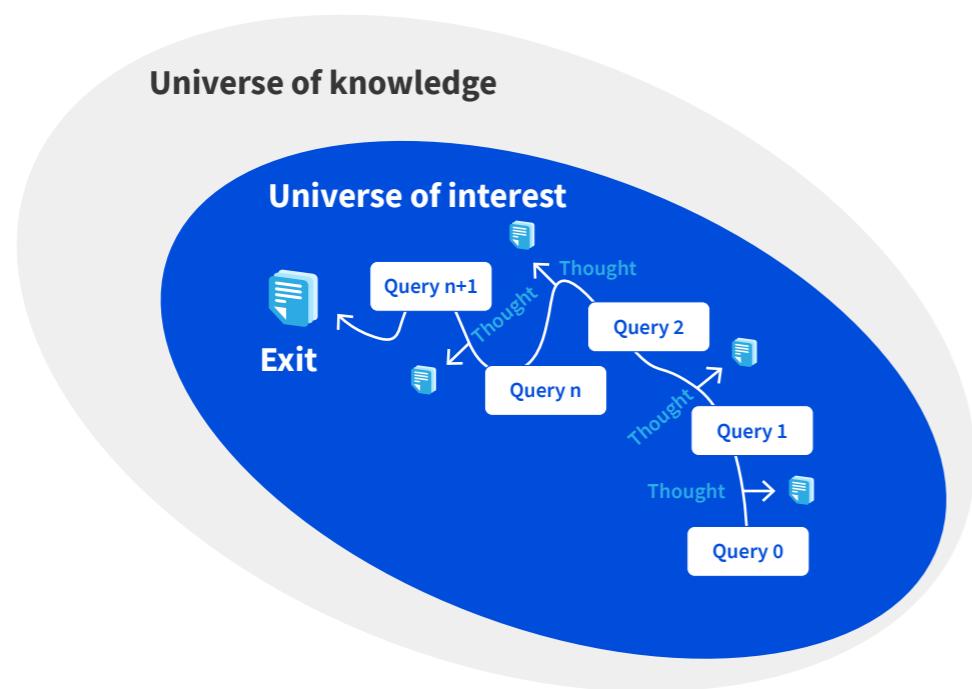


Figure 17. Berry-picking search in the context

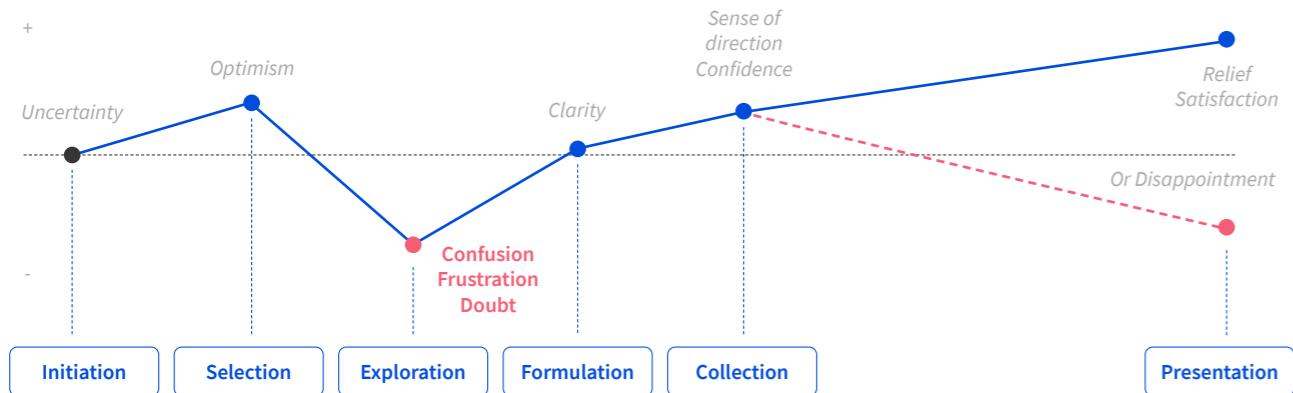
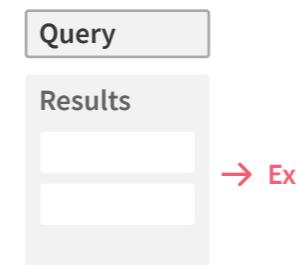


Figure 18. Stage model by Kuhlthau (Adapted by author)

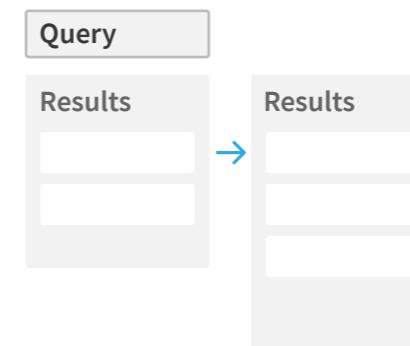
The dynamic (Berry-picking) model

Dynamic model (Figure 17), also known as berry-picking model (Bates, M. J., 1989), introduces the concept that searchers' information need adjusts as they learn from the process of interacting with the search system, requiring constant reformulation of queries and adjustments in real information need such as expanding or narrowing. Such a model has been proven by observational studies (Borgman, C. L., 1996) that search results for a goal tend to trigger new goals, and search is more than merely seeking for a set of information to a specific goal but as a process of learning and acquiring new knowledge.

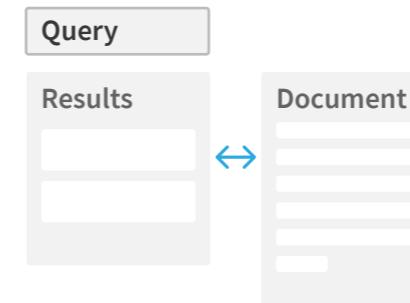
Quit



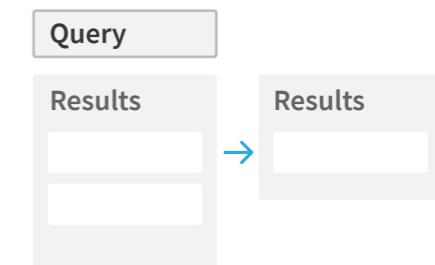
Expand



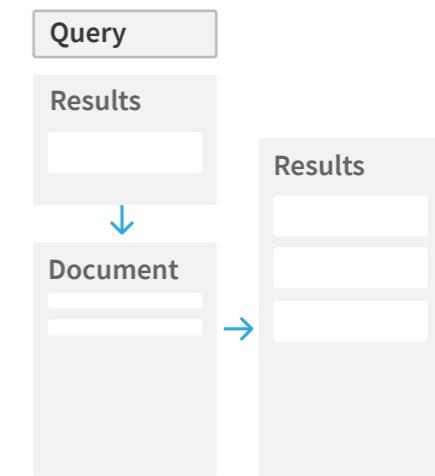
Pogo sticking



Narrow



Pearl growing



Thrashing

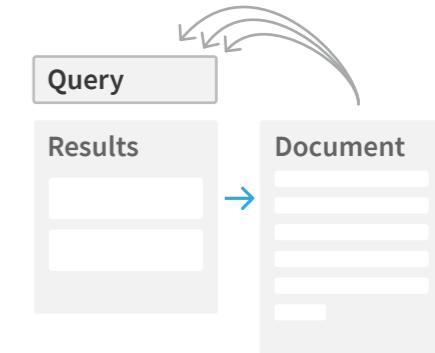


Figure 19. Search behavior patterns (image source: Search patterns 2010)

2.3.4 Seeking behaviors

Morville (2010) summarizes a series of 6 behavioral patterns (Figure 19) while a searcher is seeking information: Quit pattern, Narrow pattern, Expand pattern, Pearl growing pattern (Find one good document, then mine its content and metadata for query terms and leads), Pogo Sticking pattern (repetitive bouncing between the SERP and individual results), and Thrashing pattern (a

design flaw resides in users' heads in the form of the anchoring bias).

Such behaviors can be caused by searchers' factors as well as the design of the search system. For Quit, Narrow, Expand, and Pearl growing, they are timeless because they heavily relate to one's cognitive activities. In contrast, Pogo Sticking and thrashing appear to be anti-patterns produced by poorly designed search systems.

Animal Foraging		Information Foraging	
	Food	Goal	Information
	A site containing one or more potential sources of food	Patch	A website (or other source of information)
	Search for food	Forage	Search for Information
	The animal's assessment of how likely it is that a given patch will provide food	Scent	How promising a potential source of information appears to the user
	The totality of food types that an animal may consider in order to satisfy hunger	Diet	The totality of information sources that an animal may consider in order to satisfy an information need

Figure 20. Analogies between information foraging and animal foraging (image source: NN group, adapted by author) <https://www.nngroup.com/articles/information-foraging/>

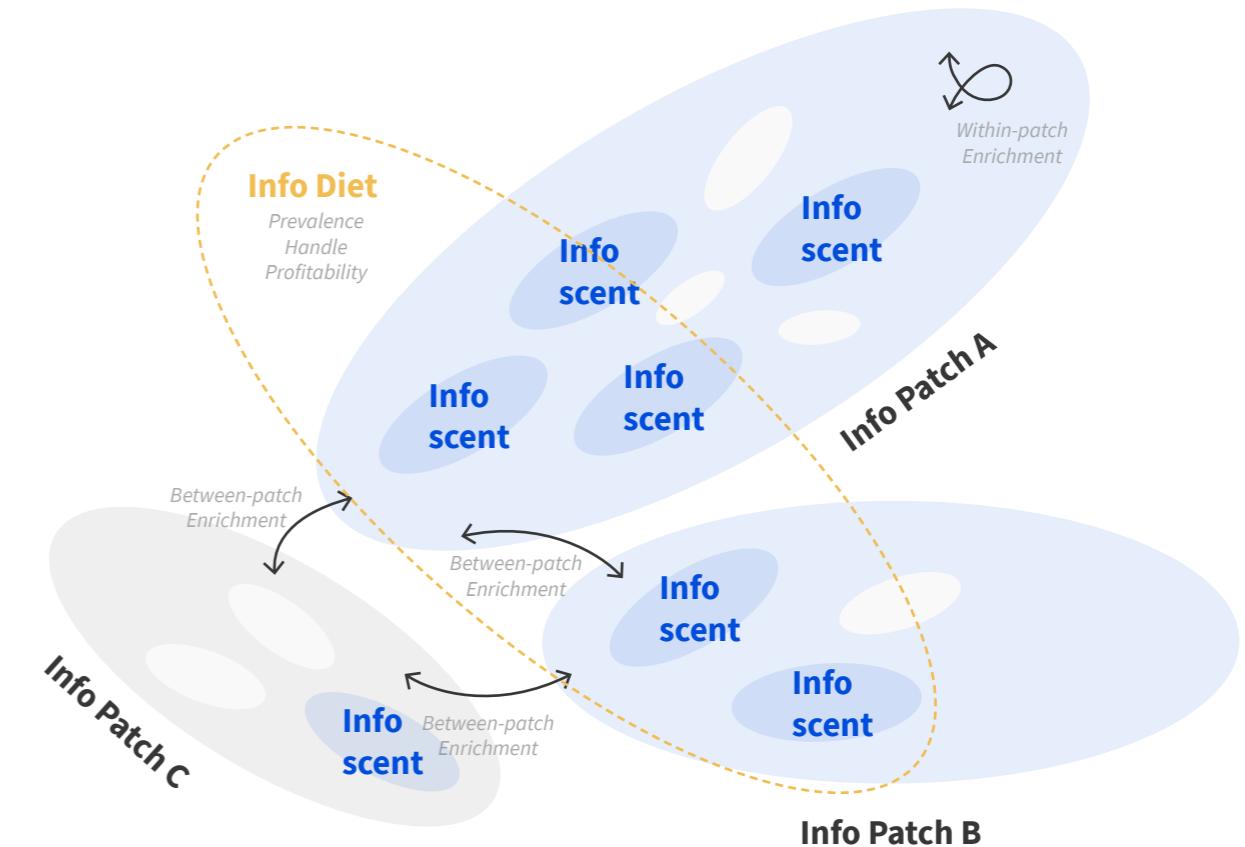


Figure 21. Information foraging theory (drawn by author)

2.3.5 Information foraging as a search strategy

Information foraging theory is analogous to evolutionary ecological explanations of food foraging strategies in anthropology and behavioral ecology (Figure 20). It is the fundamental theory of how users navigate on the web to satisfy an information need (Budiu, R., 2019). In this theory, three aspects influence the spared efforts and gains through information seeking: information patch, Information scent, and Information diet.

Search is an interplay of analytical and interactive problem solving strategies.

(Marchionini et al. 2000)

Information patch

Information patch concerns the environment where a forager seeks for information, and the foraging activity in information patches reveals two approaches: Between-patch foraging and Within-patch foraging.

There are two strategies of enrichment and exploitation that could improve the foraging results. By definition, enrichment means that the information forager can often mold the environment to fit the available strategies (e.g., minimize the between patch foraging cost). In contrast, exploitation allows a forager to modify the environment to improve within-patch foraging results (e.g., query refinement, filtering). From the perspective of User Experience (UX), however, such enrichments could cost users extra interaction and take up a more extensive cognitive load.

Information scent

Information scent is the (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues such as the link or graphical icon. A similar concept was named Information residue (Furnas, G. W., 1997), which refers to imperfect information at intermediate locations used by the forager to decide on paths through a database to target information. Information scent offers users the perception of information value and influences the following operations. Hence, Information scent should explicitly describe what the users will find at the destination.

Information diet

In facing multiple information choices, to maximize the rate of gain of information relevant to users'

tasks of information need. Prevalence, efforts to handle, and profitability are essential criteria to decide whether the forager should spare efforts to gain and consume the information. Information scent provides a sense of such criteria.

Figure 21 visualized the relationship among all three aspects involved in information foraging theory. A searcher seeks for information within or between patches, and chooses to diet certain pieces of information based on the cognitive evaluation on the gains and pays.

2.4 Relevance feedback foundations

This section strives to define and interpret relevance feedback by breaking down its constitutional factors and looking into its usage in practice and research.

2.4.1 Relevance and dimensions

Relevance

In information retrieval (IR), relevance is an important metric to measure the success of an IR system. However, measuring relevance is complicated because of the involvement of searchers' subjective judgment in contexts. Therefore, researchers split relevance as system relevance and user-based relevance (Hjørland, 2010).

Many IR research studies on relevance are focused solely on topical relevance (as the aforementioned system relevance) but ignore users' perceptions of the usefulness of search results. (Jiang, 2017). However, some considered searchers' unarticulated information needs and proposed multidimensional relevance judgment (e.g., novelty, understandability, credibility, readability, effort, freshness). Moreover, relevance is dynamic, and it changes as the information seeking process (Taylor, 2012).

Information resources

Relevance = {Surrogate, Document, Information}
Surrogate < Document < Information

Representation of the user's problem/need

Presentation = {Real Information need, Perceived information need, Request, Query}
Query < Request < Perceived information need < Real Information need

Four dimensions of relevance

Mizzaro (1998) studied the meaning of relevance and proposed a framework of four relevance dimensions: information resources, representation of the user's problem/need, time, and components.

Each dimension comprises a few elements with different weights of impact. As to Information resources, it consists of surrogate, document, and information. The relevance is the sum of all elements, and the impact increases as the stated order. Users' needs contain four elements representing the need in different levels of abstraction, and the relevance decreases as the degree of abstraction goes down.

Time has an impact on relevance, as other variables would change over time (e.g., the change of information needs). Lastly, all other elements that influence relevance were grouped as components. Mizzaro listed context, topic, and task as an example. The more components it could cover, the higher relevance it will be to a user.

Time

Overtime, the perceived information need might change, following the change of request and queries

Components

Components include Topic, Task, and Context. The more components it covers, the higher the relevance

2.4.2 Feedback

The discussion of feedback dates back to antiquity, and there are mainly four perspectives into feedback: feedback perspective, Cybernetic perspective, social science perspective and information feedback perspective. In IR, feedback relates to a task and problem-at-hand, and a users' cognitive structure and affective intentions.

Feedback loop in IR

The feedback loop is an interaction which consists of 1) a query, 2) a process to obtain a text as a response to a query, 3) the text of the response, 4) an interpretation by an interpreter on the appropriateness of the text to whatever contextual (cognitive, affective or situational) variables, and then 5) an action to modify in some way the query or the retrieval process.

Feedback types

Content relevance feedback

User query followed by an IR system output of retrieved items then judged by the user for relevance followed by a query or reformulation.

Term relevant feedback

User query followed by an IR system output of retrieved items and user selection of a new search term(s) form the retrieved output used in a subsequent query.

Magnitude feedback

User query followed by a judgement based on the size of the output from a query that affects the next query.

Tactical review feedback

User input followed by a strategy related judgement to display the search strategy history influencing the subsequent query.

Term review feedback

User input followed by a strategy related judgement to display terms in the inverted file influencing the subsequent query.

2.4.3 Relevance feedback

Relevance feedback refers to an interaction cycle in which the user reads retrieved documents and marks those that appear to be relevant, and the system then uses features derived from these selected relevant documents to revise the original query.

(Ruthven and Lalmas, 2003)

Relevance feedback is to involve the user in the retrieval process to improve the final result set. In particular, the user gives feedback on the relevance of documents in an initial set of results. It is mainly a recall enhancing strategy in traditional IR systems. In web search, relevance feedback has been used to boost personalization individually and collectively (Hearst, 2009).

Relevance feedback in AI systems

While in modern AI systems, machine learning algorithms consider relevance feedback as training data. Figure 22 (Schnabel, 2019) shows a high-level structure of a typical recommender system where the graphical user interface (UI) relays back the generation or collection of feedback data.

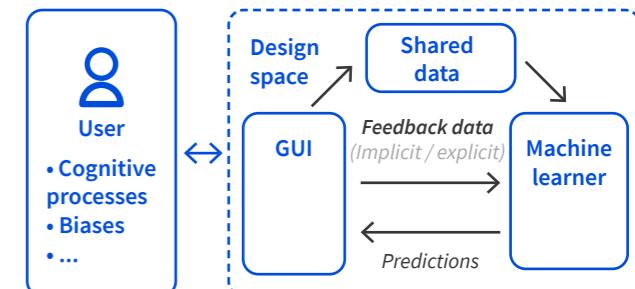


Figure 22. Typical recommender system (image source: Schnabel, 2019; adapted by author)

2.4.4 Implicit feedback vs Explicit feedback

In general, relevance feedback could be divided into two types: implicit feedback and explicit feedback.

Implicit feedback

Implicit feedback refers to the actions and behaviors users have with computational systems (Hu et al., 2008). It is readily available and unobtrusive as long as actions are being produced. It has been a significant driver in developing better information systems. In many cases, implicit feedback (e.g., dwell time and click-through data) has a strong relation to users' interests or preferences. For instance, Joachims et al. (2017) found that click-throughs have a decent correlation to the explicit judgment of relevance. Moreover, implicit feedback could be derived from various user behaviors. Kelly (2003) classified (table fixture) user behaviors that can be used for implicit feedback.

However, implicit is indirect, inherently noisy (Hu et al., 2008), and it is guesswork to infer users' interests or preferences. Plus, it is not always reliable. One study (Quiroga, 2002) showed that implicit feedback produces inferior results than explicit feedback. Another study that collects in situ explicit feedback on a web search engine suggests a different interpretation of behavior signals. The dwell time threshold between negative and positive in situ feedback is 87 seconds, longer than the more common heuristic of 30 seconds.

Explicit feedback

Explicit feedback (e.g. Netflix ratings) refers to the feedback users give directly to a computational system, meaning that it requires searchers to provide feedback by interacting with information systems explicitly. Because of its directness, explicit feedback is much more reliable than implicit feedback (Hu et al., 2008). For example, one study (Lagun, 2013) on using explicit feedback to improve search results in a location-sensitive context showed that users interact with it frequently and did lead to more efficient

searching tasks.

Recommender systems have extensively explored explicit feedback to recommend more relevant information to users. However, it has been challenging to obtain sufficient and representative feedback from a population of users. The cognitive effort can partially explain this reluctance to provide explicit feedback. Besides, a poor understanding of why providing feedback might be useful, how it should be used in the search (Ruthven and Lalmas, 2003), as well as the lack of incentives to encourage users to provide feedback (Croft et al., 2001). Moreover, UI design of feedback collection (Dooms et al., 2011) could hinder users from providing explicit feedback.

A previous study of an online music recommendation service (Jawaheer, 2010) reveals that the provision of explicit feedback from users decreases over time. Furthermore, the feedback request might be missed or ignored by many users, and potentially brings inconvenience and disrupts the search process (Kim, 2016).

Minimum Scope			
	Segment	Object	Class
Behaviour category	View Listen Scroll Find Query	Select	Browse
	Print	Bookmark Save Delete Purchase Email	Subscribe
	Copy-and-paste Quote	Forward Reply Link Cite	
	Mark up	Rate Publish	Organize
	Type Edit	Author	

Table 1. Classification of behaviors that can be used for implicit feedback.

2.5 Theories on engagement

Due to the inherent drawbacks of implicit feedback, it is insufficient for myTomorrows treatment search. Thus, collecting explicit feedback is a more appropriate approach to tackle the challenge. However, how to engage users to provide relevance feedback actively remains unknown.

2.5.1 Engagement theory

Engagement theory was first related to seeking answers to the challenge. Research in engagement with information systems has defined engagement as a category of user experience characterized by attributes of challenge(O'Brien & Toms, 2008), positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control.

On top of the study, another research recognized another two characteristics in user engagement: reputation, trust and expectation, and user

context (Attfield et al., 2011).

O'Brien & Toms conducted the explorative study with a model (Figure 23) of engagement, which is an engagement process of four stages: Point of engagement, Period of engagement, Disengagement, and Re-engagement. Different attributes influence user engagement of each stage. However, they also identified cases where users would be staying out of the engagement process (Nonengagement). For instance, users don't allow themselves to be engaged because they don't feel that they have enough time to take in the experience.

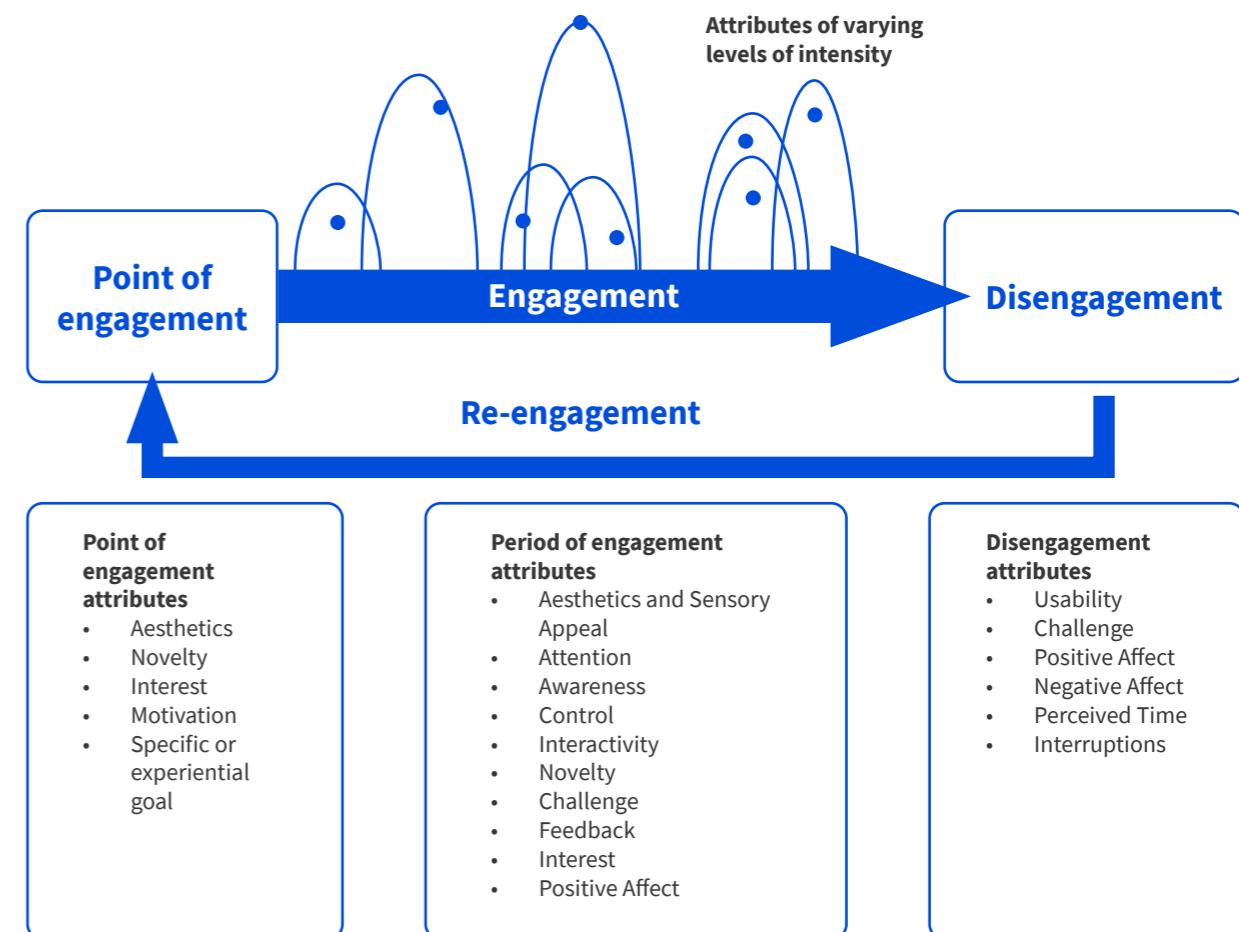


Figure 23. Model of engagement and its attributes (image source: O'Brien & Toms, 2008)

2.5.2 Nudging theory

A nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not.

(Thaler and Sunstein, 2009)

When users are facing the choices of providing feedback or not, they might need a nudge. Thaler and Sunstein (2009) proposed the nudge theory in behavioral economics and outlined the choice architecture containing six principles: **Incentives, understand mappings, defaults, give feedback, expect error, and structure complex choices.**

As an extension to the previous work, researchers (Weinmann et al., 2016) studied nudge theory in online environments (digital nudging) where user-interface design elements are used to guide people into behaving in particular ways. They (2018) also synthesized a cycled model of designing digital nudges. This cycle model includes four steps: defining the goal, understand the users, design the nudge, and test the nudge.

2.5.3 Persuasive theory and behavioral change

A gentle nudge may not be able to fulfil the job of asking users to provide feedback since for nudging to work, users might already have the tendency of doing and it is a matter of making a better choice. Hence, persuasion design and behavior change were looked into to decode how users would behave in a certain way.

A behavior model of persuasive technology

Fogg (2007) proposed a behavior model (FBM, Figure 24) for persuasive design, and it argues that behavior is associated with 3 factors: motivation, ability, and triggers. To perform a target behavior, one must be sufficiently motivated, has the ability to perform the behavior, and be properly triggered.

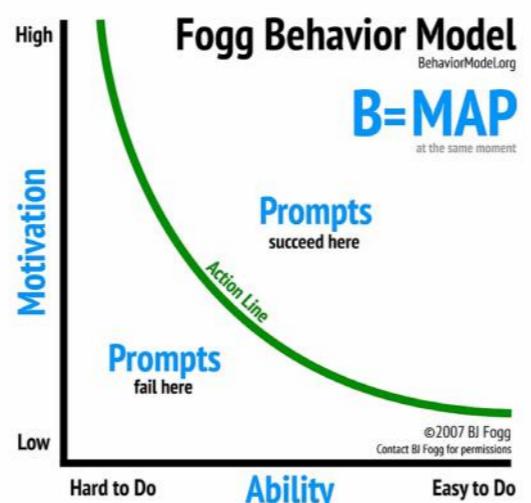


Figure 24. FBM a behavior model for persuasive design (B.J.Fogg, 2007)

Motivation

There are 3 types of motivations: Pleasure/Pain, Hope/Fear, Social acceptance Rejection. They are distinct from each other but they are all linked fundamentally with human sensory and evolutions.

Pleasure and pain: Such motivators are immediate or nearly so, people are responding to what's happening in the moment.

Hope and Fear: This dimension is characterized by anticipation of an outcome. It is more powerful at times than pleasure/pain, but not necessarily more motivating than that.

Social acceptance and Rejection: Much of our behaviors are controlled by social norms, and people tend to win social acceptance and avoid rejection.

Ability

To perform certain behaviors, there is always the activation threshold. There are 6 elements that determine one's ability: time, money, physical effort, brain cycles, social deviance, non-routine. Different people vary in the combinations of those elements, but what everyone shares is resisting attempts at motivation, and humans naturally love simplicity. Therefore, persuasive design relies heavily on simplicity.

Triggers (Prompt)

As people vary in ability and their motivations, there are 3 identified triggers: spark trigger, facilitator, and signal. Each serves for different combinations of varied levels of ability and motivations.

Spark trigger: one lacks motivation, a trigger should be designed in tandem with a motivational element.

Facilitator trigger: when one has high motivation but lacks ability, a facilitator aims at triggering the behavior while making it easier to do.

Signal: when one is both motivated and has the ability, and it just serves as a reminder.

Behavioral change types

In the work Behavior change support system (Oinas-kukkonen, 2010), the researcher pointed out that for behavioral change, a system should be usable and useful. He defined three types of behavior changes:

C change: users comply with the request of an information system

B change: a more enduring change and will be sustained in a long run (for this to happen, A change is needed)

A change: to influence users' attitudes rather than merely single behavior.

2.5.4 Persuasion cues

Fogg (2003) proposed computers as persuasive

social actors, and listed five primary types of social cues (Table 2) with five persuasive principles.

Cue	Principle
Physical	Principle of attractiveness A computing technology that is visually attractive to target users is likely to be more persuasive as well.
Psychological	Principle of similarity People are more readily persuaded by computing technology products that are similar to themselves in some way.
Language	Principle of praise By offering praise, via words, images, symbols, or sounds, computing technology can lead users to be more open to persuasion.
Social dynamics	Principle of Reciprocity People will feel the need to reciprocate when computing technology has done a favor for them.
Social roles	Principle of Authority Computing technology that assumes roles of authority will have enhanced powers of persuasion.

Table 2. Five persuasive principles

2.5.5 Theories in motivation

Theories of engagement, nudging, and persuasion all mentioned motivation, thus, this might be a key to tackle the challenge of collecting explicit relevance feedback.

Types of motivation

Motivation can be described as a process to release, control, and maintain physical and mental activities. In general, there are two types of motivations: intrinsic motivation and extrinsic motivation. (Deci and Ryan, 1985). Intrinsic motivation exists if an individual is activated because of seeking the fulfillment generated by the activity (e.g. acting just for fun). In contrast, Extrinsic motivation is just an instrument for achieving a particular desired outcome (e.g. acting for money or avoiding sanctions).

Studies in crowdsourcing (Mason, 2009) have found that extrinsic motivation factors such as monetary reward merely increase workers' willingness to accept a task or speed of completion, but does not improve the quality of work. Another study (Rogstadius et al., 2011) showed that intrinsic motivation (e.g. framing a task for helping others) succeeded in improving the output quality. Deci (1971) found out that extrinsic and intrinsic motivations relate to each other. For instance, intrinsic motivation decreases when money is used as an external reward, while verbal reinforcement and positive feedback tend to increase intrinsic motivation.

A motivation model in crowdsourcing

Due to the nature of relevance feedback provision, it could be regarded as an extension of crowdsourcing, where motivation is extensively studied. Researchers (Kaufmann et al., 2011) have synthesized and proposed a model (Figure 25) for workers' motivation in crowdsourcing. The model complies with what Deci and Ryan have discovered, identifying two categories under intrinsic motivation, three categories of extrinsic

motivation.

Intrinsic motivation

Enjoyment based motivation contains factors that lead to the sensation of "fun" that might be perceived by the workers.

Community based motivation covers the acting of workers guided by the platform community

Extrinsic motivation

Immediate payoffs cover all kinds of immediately received compensations (such as payment) for the work on crowdsourcing tasks

Delayed payoffs address all kind of benefits that can be used strategically to generate future material advantages

Social motivation is the extrinsic counterpart of intrinsic motivation by community identification. It covers socially motivated extrinsic motivation out of values, norms, and obligations from outside a platform community as well as indirect feedback from the job and the need for social contact.

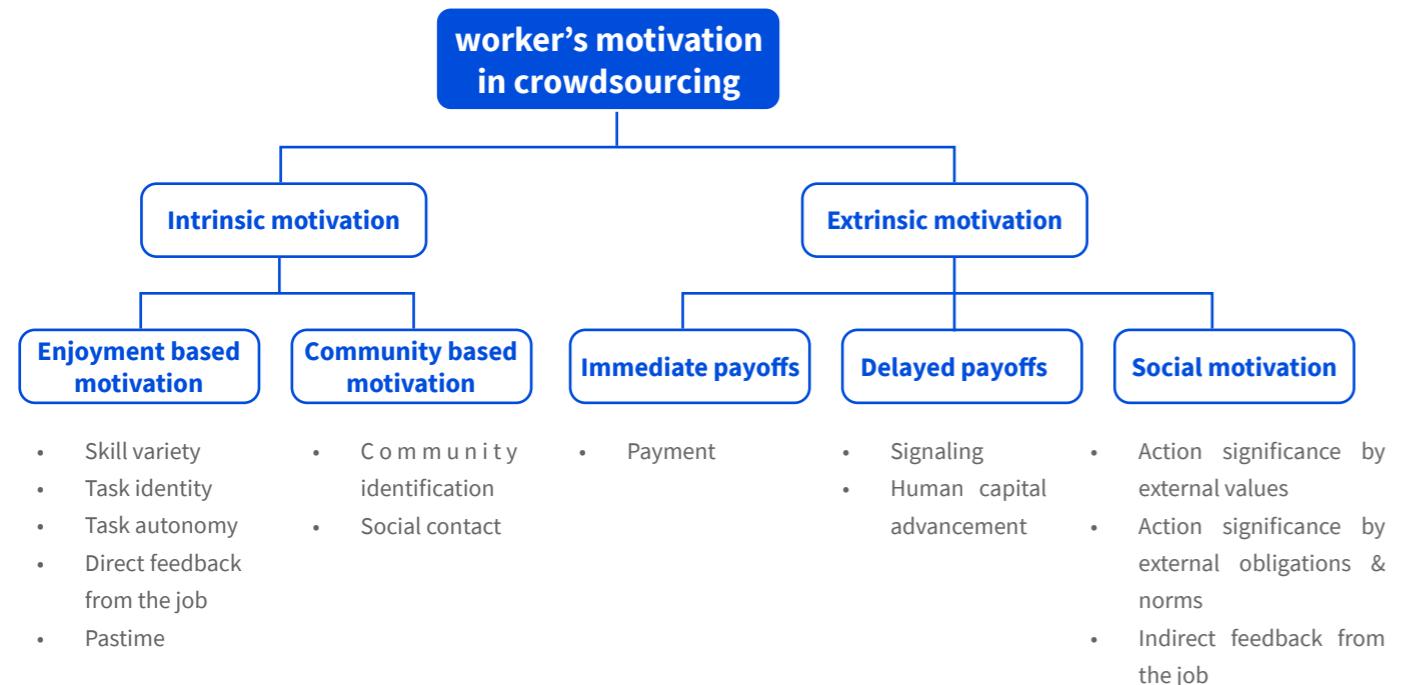


Figure 25. Figure of motivation constructs (source: Kaufmann et al., 2011, adapted by author)

Motivation in designing AI systems

Similar work has been done by Google (People + AI Guidebook), which identifies five canonical reasons together with their pros and cons: Material rewards, Symbolic rewards, Personal utility, Altruism, and Intrinsic motivation.

Material rewards

Cash payments are highly motivating. Mechanical Turk is an example of this type of reward for feedback at scale.

Pros:

- A direct solution to increase feedback
- May increase the volume of feedback

Cons:

- Costly to run over time
- May devalue intrinsic motivations
- Biases for a subset of users
- May decrease feedback quality

Symbolic rewards

These can include status attainment, such as virtual badges, social proof and group status by projecting a self image to a community, and social capital, such as a reputation as an expert.

Pros:

- Low to no cost

Cons:

- Relies on users caring about how they're perceived
- Creates power imbalances in the community
- May inhibit intrinsic motivation

Personal utility

These include "quantified self" experiences including allowing users to track their progress, bookmark things for later, and explicitly training a personalized AI model — like a recommendation engine — for more relevant output later on.

Pros:

- No network effects or community necessary

to begin

Cons:

- Privacy does not support community development
- May inhibit intrinsic motivation

Altruism

Altruistic motivations can include community building and helping other people make decisions, such as leaving a product review, as well as trying to increase fairness, like giving a conflicting opinion by disagreeing with a particular product review.

Pros:

- Potential for more honest feedback based on a desire to help

Cons:

- Social desirability biases may lead to extremes in feedback content
- Decrease in contributions if the opinion is already represented
- Altruism levels may vary across cultures or groups

Intrinsic motivation

Intrinsic motivation is the internal fulfillment people get from the act of expressing themselves. This includes direct enjoyment from giving feedback, the ability to vent and express opinions, and the enjoyment of community participation.

Pros:

- No network effects or community needed to start
- People like to do things they enjoy

Cons:

- Social desirability biases may lead to extremes in feedback content

2.6 Perceived trust in search

Perceived trust in Health websites have a high priority in user engagement (Wortham, 2009). Besides, without human trust in a system, efficiency, productivity, and user experience will not be maximized. Without trust, users will seek to achieve the tasks in other ways (Muir & Moray, 1996). To understand what trust is and what comprises it, hence, knowing how design can interfere and enhance users' perceived trust in the search.

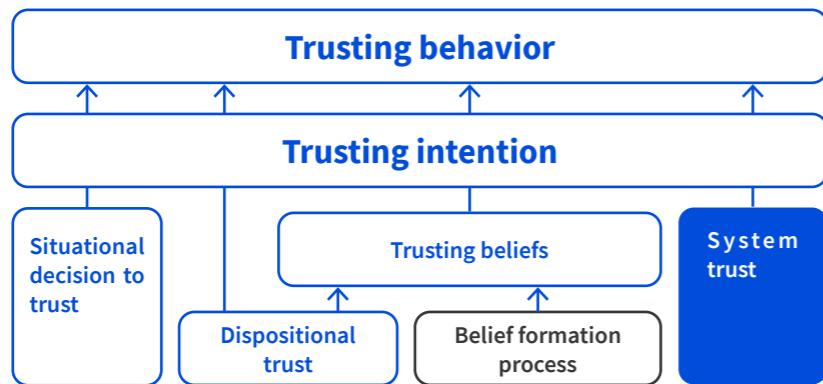


Figure 26. McKnight trusting model that involves 6 dimensions of trust construct (image source: McKnight, adapted by author)

2.6.1 What is trust?

Trust is based on an individual's theory as to how another person will perform on some future occasion, as a function of that target person's current and previous claims, either implicit or explicit, as to how they will behave.

(Good, 1988)

Trust is a broad, fuzzy, yet difficult-defining concept. Previous studies stem mostly from the subjects of sociology, psychology, and philosophy. In general, studies have categorized trust as dispositional trust, learned trust, and situational trust.

Trust is subjective, and for trust to occur, risk (March, 1995) and uncertainty must be perceived by the trusting party. Trust resides in one individual

and is something to which that individual alone has direct access. Moreover, it requires another individual, or to this specific context, an artifact of technology as a stimulus.

2.6.2 Trusting dimensions and models

6 dimensions of trust construct proposed by McKnight (figure 2.1.4) in 1998 is widely adopted in the economic and organizational studies of trust, and it serves as the foundation of many recent research on trusting models in electronic environments (Chopra, K., & Wallace, W. A. 2003) or information systems such as e-commerce.

Apart from research on building blocks of trust and designing universal trusting models, it is also widely studied around trust on content presented online (Gil, Y., & Artz, D. 2006) and design guidelines of increasing trustworthiness in information systems such as Fogg (2001) and NN group (2016).

6 Trusting dimensions

Dispositional trust: a generalized expectation about trustworthiness of others.

Trusting beliefs: the extent to which one believes something/someone is trustworthy in the situation.

System trust: or Institution-based trust, one believes that proper impersonal structures are in place to enable one to anticipate a successful future endeavor.

Situational decision to trust: the extent to which one intends to depend on a non-specific other party in a given situation.

Trusting intention: or Willingness to trust, the extent to which one is willing to depend on other in a given situation with a feeling of relative security.

Trusting behavior: is the extent to which one person voluntarily depends on another person in a specific situation with a feeling of relative

security, even though negative consequences are possible.

2.6.3 System trust in search

Trust is a subjective form of experience rooted in an individual's dispositional traits, and surrounding factors influence it. The scope of this project will focus on increasing trustworthiness through system trust as to that of the model mentioned earlier. Impersonal factors of building system trust include the technology artifact, namely the Information system and its Interface (Figure 29).

Factors impacting trust in search Interface include visual (static trust), Interaction(dynamic trust), and information. Behind the manifestation, users' perception of the information system via interactions with the user interface is vital because the deriving of information one is seeking relies on it. Apart from the mechanism of information provision, technology bias is another critical factor affecting trustworthiness needing addressing through the interference of design.

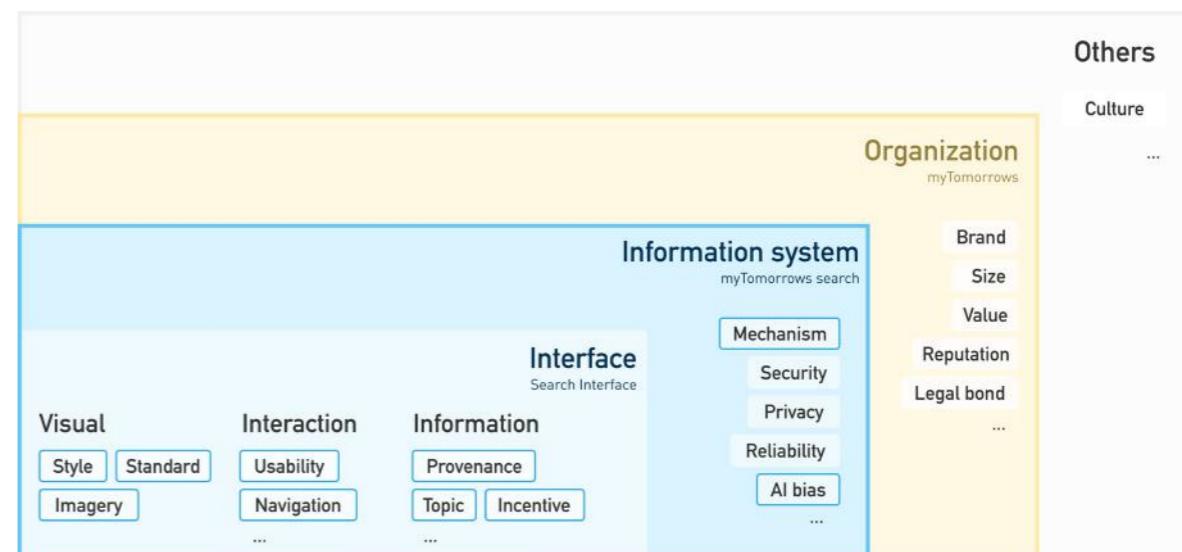


Figure 27. Scope of the trust factors in this project (Drawn by author)

Key takeaways chapter 02

2.1 Search foundations

Search, or information retrieval, has a long history in information science, and its powering technology has been transformed through four main stages. Nowadays, search engines are more than just matching identical search queries from the information. Rather, it can support high-level semantic search in the connected web of information (semantic web) with the help of AI technologies (e.g. Machine learning for recommendations).

2.2 The design of search

Research around the design of search (IR) systems has identified five elements (user, interface, engine, content, creator) in search and summarized ten design patterns (Autocomplete, Best first, Federated search, Facet search, Advanced search, Personalization, Pagination, Structured results, Actionable results, and Unified Discovery) and proposed design guidelines for search interface.

2.3 Search as information seeking

Search or seeking information is a dynamic process that requires constant examination of the information retrieved to specific queries. In general, there are three types of intentions for one to search: navigational, informational, and transactional. For this project, the purposes of users in myTomorrows search will be informational or navigational.

Nowadays, search interface design focuses on supporting the information seeking process by enhancing users to (re)formulate search queries and evaluate results. Search is a complex, repetitive, and dynamic process in which users' information needs evolve as they perform searches. The process could be generally divided into three stages of search preparation, interaction

with search system and results collection, and actions on search results. Consequently, assisting users to quickly identify the value of information and reducing the cost of seeking is the core strategy for search systems, as indicated in information foraging theory.

2.4 Relevance feedback foundations

Traditionally, relevance in IR is an important metric to measure the success of an IR system, but it solely focuses on topical relevance, ignoring the human factors. However, relevance is subjective and difficult to measure. Most recent studies around relevance have put more attention to human factors, and four dimensions of relevance were identified: information resources, representation of the user's problem/need, time, and components.

Relevance feedback in AI systems (e.g. recommender system) is often used as the data for training the machine learner with Machine learning algorithms. Generally, relevance feedback fed in AI systems can be split into implicit and explicit feedback. The former has been the main drive in many information systems as its unobtrusive nature and effectiveness in practice. In contrast, the latter has shown a strong performance in many recommender systems, but it requires extra cognitive burden from end-users. Previous research has seen a drop in explicit feedback provision from users and even reluctance to doing this. Thus, the main issue lies in engaging and motivating users to provide explicit feedback.

2.5 Theories on engagement

To engage users to provide relevance feedback is essential, and multiple factors influence an engaging experience. Theories around user engagement have identified attributes of challenge, positive affect, durability, aesthetic

and sensory appeal, attention, feedback, variety/novelty, interactivity, perceived user control, reputation, trust and expectation, and user context. All those factors are crucial when designing for the search interface. Moreover, The motivation of users should be the focus while designing feedback collection interactions.

2.6 Perceived trust in search

The perceived trust is one crucial factor for user engagement. Also, it is vital in such a sensitive context of this project as the information. However, trust is a broad topic. In theory, for trust to occur, risk and uncertainty should be perceived by the trustee. According to the trusting model proposed by McKnight, there are six trusting dimensions, including impersonal factors such as system trust. Technology and its provider influence each other, and together, they form partial trust in technology (or system). The scope of trust in this project will be focused on system trust, and to be more specific, the information system (Mechanism, AI bias) and its user interface (Visual, Interaction, Information).

03

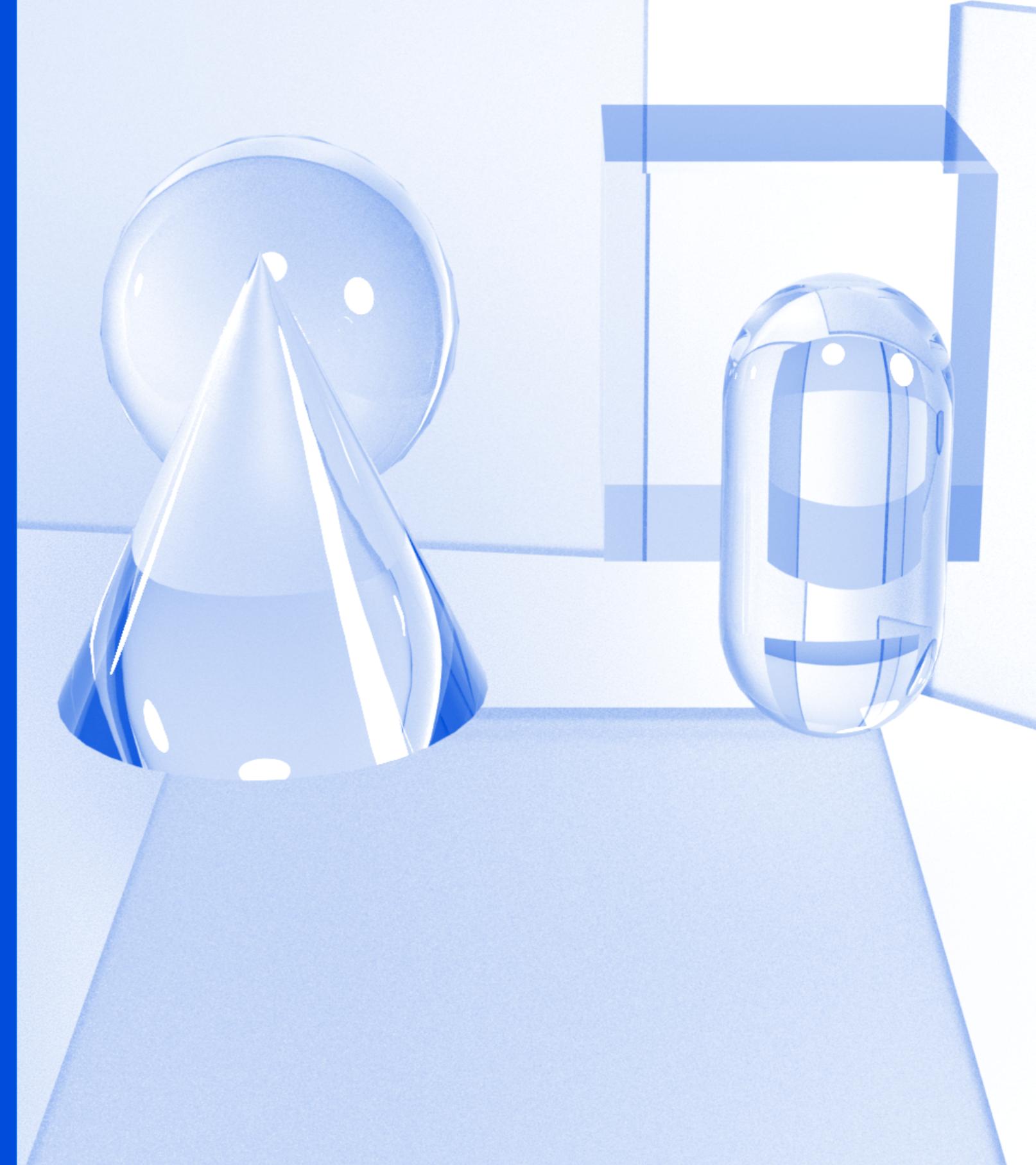
The current, exploratory research



This chapter aims at collecting data and insights from the real world and end-users (healthcare professionals), to fill the gap that the literature review can not mitigate. The exploratory research begins with understanding the current product and collection of Search future visions from myTomorrows and industry. Most importantly, it approaches the end-users to discover the current challenges they face for eliciting design opportunities and their rationale for relevance judgments. Lastly, explicit feedback collection in practice was looked into for design inspirations.

Chapter overview

- 3.1 Product analysis
- 3.2 Future visions on search
- 3.3 User research (healthcare professionals)
- 3.4 Building blocks of feedback
- 3.5 Feedback collection in practice



3.1 Product analysis

Product analysis of myTomorrows search dives into the fundamentals by decomposing the buildup and inquiring underpinnings, as well as evaluating search experience in use. The objective is to understand how the Search was constructed and benchmark the user experience. Two activities were carried out in this stage: product decomposition and Heuristic Evaluation.

3.1.1 Search decomposition

AI-powered search architecture

myTomorrows treatment Search (Figure 29) builds on AI technologies. The knowledge graph plays a vital role, and it incorporates a broad spectrum of medical data, for instance, WikiData and Unified Medical Language System data. Unlike most web search engines that crawl metadata from various sources (p.27), myTomorrows Search returns only text-based data and relies heavily on the knowledge graph to which Clinical Trial (CT) and Expanded Access Program (EAP) data are affiliated.

CT/EAP source

There are 18 primary clinical trial registers (Figure

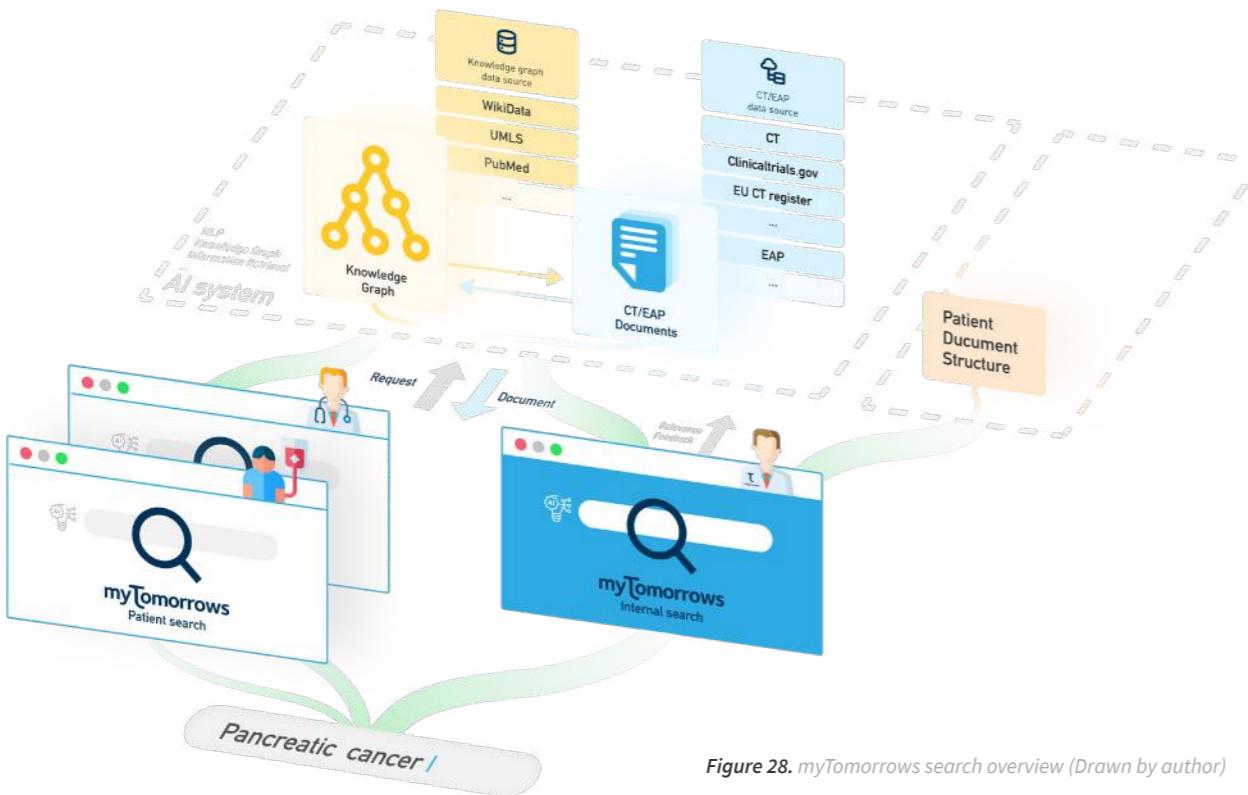


Figure 28. myTomorrows search overview (Drawn by author)

CT/EAP data structure

Different CT/EAP databases adopt diverse data structuring standards, but they partially share common information. Since Clinicaltrials.gov is the largest CT register, its CT/EAP data structure was looked into as the basis of understanding what information a CT/EAP comprises. Table 3 shows a simplified data structure of information in different information resources (p. 40).

Surrogate	Document	Information
Status	Study Description	The information of investigational treatment that is able to both meet the patient treatment needs and is convincing to HCPs base on their knowledge/experience
Study Title	Study Design	
Conditions	Arms and Interventions	
Intervention	Eligibility Criteria	
Phase	Contacts and Locations	
	More information	

Table 3. Information structure of CT/EAP on Clinicaltrials.gov

Ordering but NO ranking

Due to regulatory compliances and the avoidance of potential biases in data, myTomorrows search does not rank search results in the search result page (SERP). Nevertheless, Morville (2003) identified that users would perceive higher positioned search results more relevant to his/her information needs.

At present, data types (CT/EAP) and phases of CTs are the underpinnings of ordering search results. CTs would be listed prior to EAPs because of regulatory reasons (p. 20). Within CTs, later phases come before earlier phases as interventions being studied in phase 4 would be less experimental and supported with more efficacy data.

Information architecture

myTomorrows external search comprises two portals (grey coded) respectively for HCP and patient. The information architecture (Figure 30) of the HCP search consists of 6 building blocks (colored blue). myT treatment Search (LOGO), Guidance information, Search criteria, Filter, CT/EAP, Access. The yellow coded parts stand for linkage to external resources.

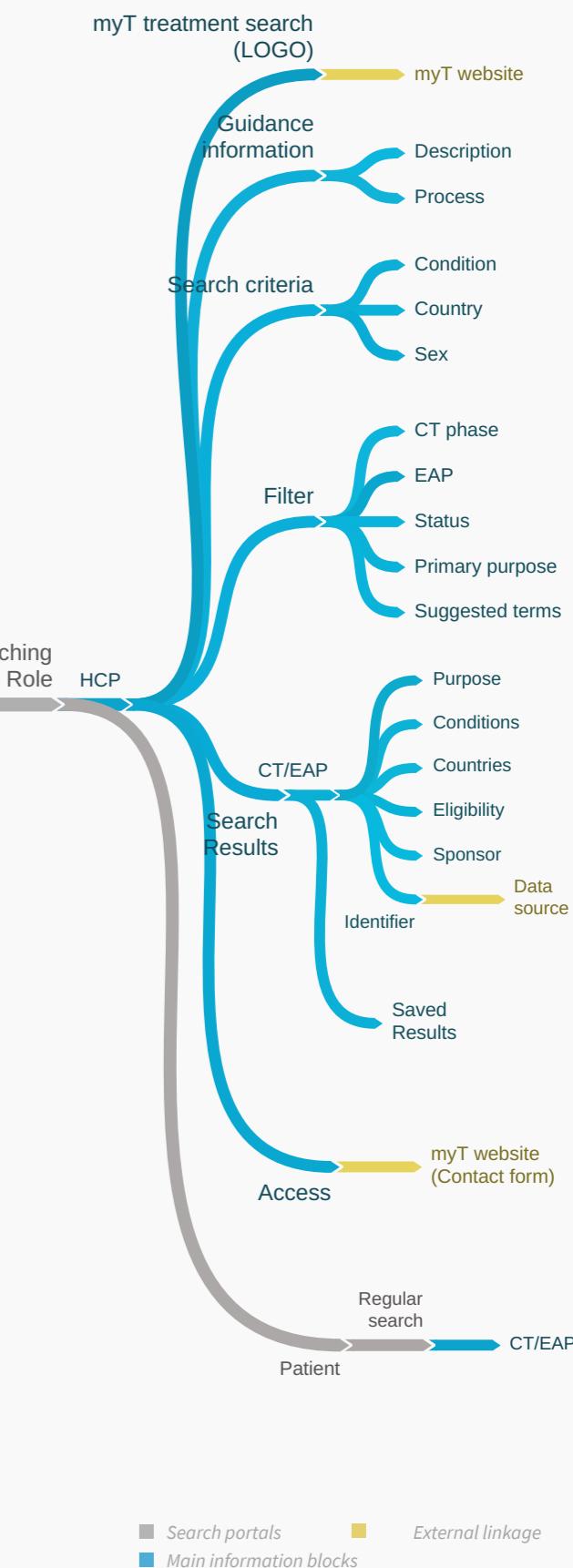


Figure 30. The information architecture of current myTomorrow external search (HCP portal)

3.1.2 Heuristic evaluation

Heuristic evaluation on the external search was conducted with peers to benchmark the current experience, identify issues in UX, as well as an approach to understanding the product deeper.

Approach

The evaluation recruited twelve peers with four different educational backgrounds (Figure 31). Ten sessions were conducted, including two with two participants at the same time.

The procedure comprised three sections: context introduction, free exploration, and reflection. In the second section of free exploration, participants were asked to think aloud. The data from the study comprises notes following the ten Usability Heuristics for user interface design (Nielsen & Molich, 1990) and first impressions of evaluators.

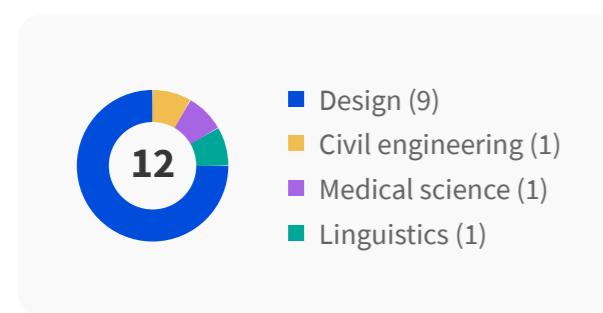


Figure 31. 12 evaluators with 4 different backgrounds

First Impression of evaluators

Most evaluators (8/12) thought the system is clean, and information is well organized. However, four evaluators felt the system is not a ready product but an internal tool or solely a prototype. Half of the evaluators pointed out that a lot of interactions and visual details do not align with their expectations. Plus, the experience gets overwhelming over time, partially because of the medical knowledge gap.

Moreover, nearly 50% of evaluators expressed their concerns that the Patient portal has a very high threshold for medical knowledge needed. One mentioned that the search is, somehow, disconnected from myTomorrows service.

Primary heuristic results

Following the ten Usability Heuristic criteria, 106 points were gathered (Table 4). Major problems of the current search fall on ‘Flexibility and efficiency of use’ with 18 identified issues, followed by ‘Consistency and standards’ (seventeen UX issues). ‘Visibility of system status’ and ‘User control and freedom’ shared the same number of issues (16). In comparison, the current system performs well in ‘Recognition rather than recall’ (3).

Visibility of system status	16
Match between system and the real world	13
User control and freedom	16
Consistency and standards	17
Error prevention	7
Recognition rather than recall	3
Flexibility and efficiency of use	18
Aesthetic and minimalist design	7
Help users recognize, diagnose, and recover from errors	5
Help and documentation	4

Table 4. unmerged issues in 10 usability heuristics

UX issue origins

Table 5 shows an overview of the origins of 51 UX issues (red) with the same issues bind together, and a sum of 18 (blue) was classified as wishes/wants. They scatter in ten UI elements of the search. The most prominent one was found in the search result list (a part of the search result page) with 12 UX issues (and one wish/want), and 8 issues came from the search box (input for condition and search criteria).

01 Search home (3)	06 Loading (2)
02 Patient/HCP search (1 + 2)	07 Result list (12 + 1)
03 Process graph (4)	08 Result content (7 + 2)
04 Search Box (8 + 4)	9 General (7 + 4)
05 Filter (5 + 1)	10 Discuss/Contact (5 + 1)

Table 5. Merged UX issues with origins

UX issue severity

Other than heuristics and identification of their origins, all issues were sorted and clustered depending on affinity and severity (Figure 32). They formed five clusters: technical bug, critical issue, major issue, minor issue, and Wishes/wants.

Most UX issues were labeled as major and minor issues, making up forty-one UX issues. In comparison, only three were regarded as critical issues, and five were technical bugs. However, technical bug and critical issues severely affect the search experience as they are closer to the backend and can easily lead to system failures. For instance, one issue belonging to the technical bug is ‘customized search query is not supported.’ In contrast, in critical issue, ‘It returns completely irrelevant search results, but was relevant to the previous search query.’

Complete results in Appendices (p.154-155)

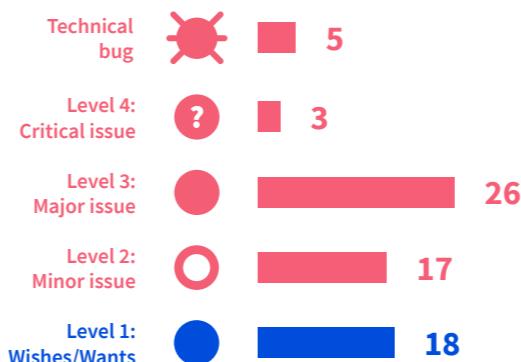


Figure 32. UX issue clusters with the number of issues in each cluster

💡 Key takeaways

- The overall experience of the current search is moderate, but with a huge room (41 major and minor issues) to improve interaction details and visual impression.
- The main focus of improvements should be offering more flexibility and increasing the efficiency of use, especially in the search result page.

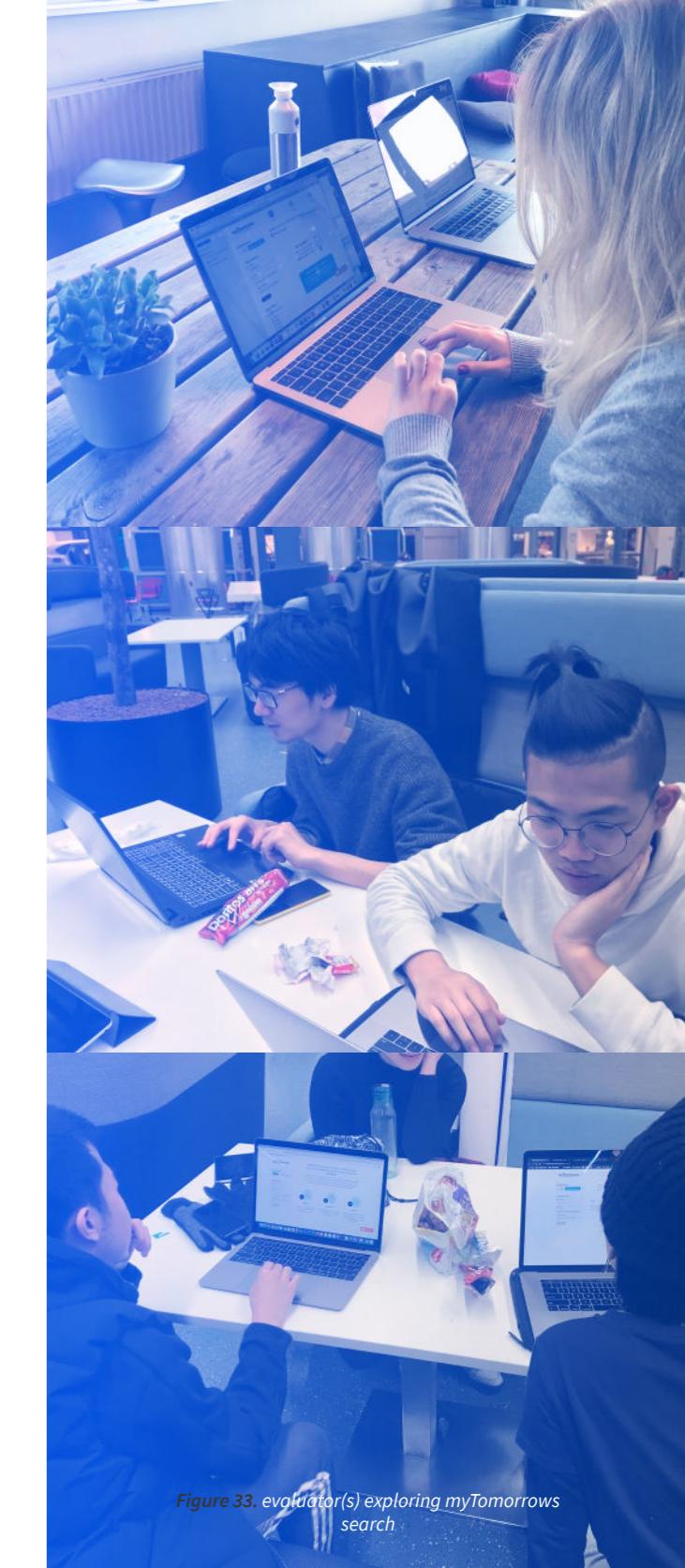


Figure 33. evaluator(s) exploring myTomorrows search

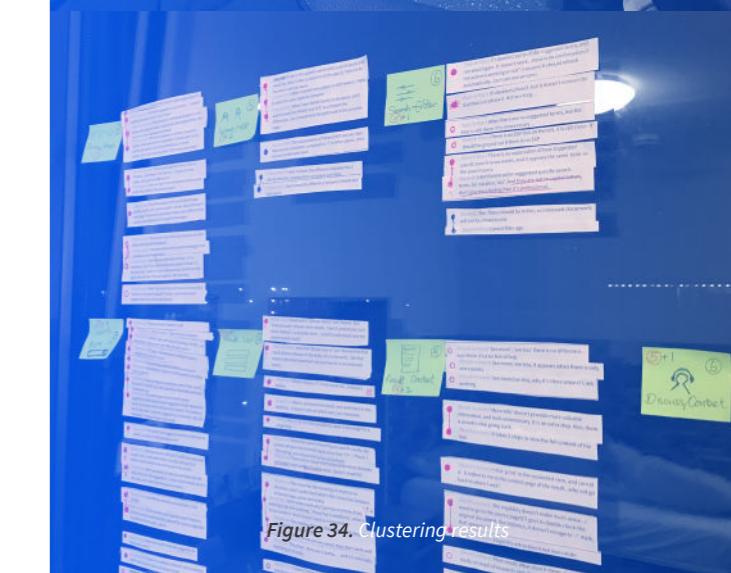


Figure 34. Clustering results



Figure 35. Thematic analysis



Figure 36. CTO (I7) drawing out service vision



3.2 Future vision

The collection of future visions on myTomorrows treatment Search aims at forming a holistic overview of the product definition for fulfilling myTomorrows' mission. Interviews with employees across different departments at myTomorrows were conducted, together with a longitudinal study on Clinicaltrials. gov modernization as a supplementary source.

3.2.1 future vision interviews

Approach

A total of seven interviewees within the company participated in the interviews. Three out of seven had professional education in healthcare-related fields, and all had more than ten months of working experience at myTomorrows.

All interviews were conducted individually with interviewees following the same protocol consisting of five questions (see below). Each session was audio-recorded and transcribed for analysis with Thematic analysis (Braun, Clarke, 2006).

Interview protocol

- How is your current experience with myTomorrow search?
- What are your future visions on myTomorrows search?
- Could you please walk me through from a patient's perspective?
- Could you please walk me through from an HCP's perspective?
- How would you pitch your visions to the boss?

Interviewees (I)

- I1, Project manager (1 yr 3 mons)
- I2, Corporate development (3 yrs)
- I3, Medical search coordinator (1 yr 6 mons)
- I4, Product manager (1 yr 7 mons)
- I5, C-Medical-O (3 yr 3 mons)
- I6, AI lead (10 mons)
- I7, CTO (3 yr 2 mons)

3.2.2 vision clusters

The results are fruitful throughout the analysis. According to the timeframe for implementation and realistic boundaries, two groups of visions were identified as near-future vision and far future vision. Besides, interviewees also described and shared the facts and concerns around the treatment search.

Near future vision

This cluster refers to visions that can be achieved within a time frame of three years with the available resources and workforce. Some of these visions exist as part of the development plan, and some sooner or later. A total of eleven themes were synthesized as below:

Complete source data

myT should have complete source data of all (17) registries and EAP programs.

Quality & rich information (Help decision-making)

The information offered in myT search should be rich and high quality, such as incorporating efficacy data to support HCPs to make decisions.

Clear actionable

myT search should communicate to HCPs what actions to take to access to products or enroll in studies.

Control & feedback on results

Users should be able to have control over irrelevant results, and feedback given should be taken into account for system optimization in a sensible way.

Result management

Previous search results should be manageable for reuse or other purposes in the future.

Search report manipulation

The search report should be brought to the search for HCPs to manipulate/interact freely (for instance, adding or removing results), instead of a mere PDF.

Insights & trends

myT search should enable HCPs to follow their fields/diseases of interest to gain insights and to learn.

Discussion & co-decision between HCPs

HCPs sometimes do not make decisions alone but may involve discussions with other HCPs, myT search could support co-decision or discussion on search report(s).

More user-friendly

myT search should be much more user-friendly than other competitors (e.g., Clinicaltrials.gov).

Patient search as an option

Patient search should be an option for those who are willing to obtain more information. However, it should help patients understand such information for better medical decisions together with their HCPs.

Lower search barrier

myT search should lower the barrier for patients to search, such as simple asking questions, while for HCPs, more abundant filters could help obtain a more relevant result list.

Far future vision

This cluster describes the visions that can be achieved beyond three years with the available resources, human resources, and the same business strategies. A total of four themes were listed, as well as one undefined theme that is marginal to the current strategy.



In-hospital search

myT search shall be integrated into HCPs' work environment (hospital) and direct access to patients' EHR data. It is expected to be the platform where HCPs go for new treatments as their routine.



Outsource search to myT

When trust is established and high enough, HCPs would outsource searches to myT and be provided with a final list of options to review.



Fully automated match

Search is expected to be fully automated when myT technical capability is high enough to auto-match with high precision. It will reduce myT workforce and increase business scalability.



A bridging platform

With multiple entry portals, the search connects patient, HCP, and pharma. myT search enables two ways of matching that allows pharma to find participants and collecting study results. Vice versa.



Other than the ones mentioned above, it was also envisioned that HCP KOL could be involved in the platform, as well as developing other business opportunities based on data and knowledge myT will have.



This cluster describes empirical knowledge around myTomorrows services, and they are divided into four themes.

Patient: Most patients have the medical knowledge gap, and not all patients are looking for treatments but information. Also, individual differences affect patients' acceptance of investigational treatments.

HCPs: Not all HCPs know EAP, and HCPs do not mind prescribing investigational treatments as long as there are convincing study data and results. Besides, HCPs sometimes listen to KOLs' opinions as a reference for prescriptions. (It lays the empirical foundation for 3.3 User research HCPs)

Technical boundaries: Adding more data sources will bring in data processing challenges due to non-standardized data. Machine translation from jargon to trustworthy layman language has a high technical threshold. However, a new regulation in the EU requires CTs to provide layman language. Currently, the knowledge graph of myT search and CT/EAP documents are highly connected.

Source data: It refers to CT/EAP data from registries. They are mostly text, multilingual, and scattered. Not all CT registries include EAPs, and data quality varies in different databases. Problems can be found in source data: outdated data, mistakes inherited from authors, and data (CT/EAP) are not connected to prior studies. Moreover, only a small fragment of the data has bond study results, but most study results would be published on other medical databases (e.g., PubMed).

Concerns

This cluster describes concerns and issues interviewees had. Three themes range from search to the services myTomorrows provides.

False impression: myTomorrows is a commercial company, and it inherently gives an impression of striving for profits. The public might perceive search results promoted or misleading.

Perceived trust: since all source data are from CT/EAP registries, it is uncertain how myT search works to retrieve data, and it might be perceived as a black box. The current data displayed on myT search is not identical to its source. Consequently, users may need to double-check and compare search results.

Current search issues: the current search has issues in the system level and the UX level (refer to 3.1.6 Heuristic evaluation). It does not provide an efficient overview of information, which lags the productivity of the internal medical team.

3.2.3 Clinicaltrial.gov Modernization

Clinicaltrials.gov (the US register) had been campaigning for modernization since 2019. As one of the most important data sources, the open-access materials from their public survey (Request for information) and Webinar were followed for illustrating a holistic overview of users' voices and the frontline of clinical studies.

Request for information

Results of Request for Information (<https://prsinfo.clinicaltrials.gov/SummaryResponsesToRFI.pdf>) covered three broad topics of website functionality, information submission, and data standards. Website functionality was the focus of this research with four subtopics:

Scope of primary use: there are three types of use scopes: 1) searching for a wide range, 2) a narrow range, 3) both. In most cases, respondents seek only for a narrow range for certain diseases, but the other two cases are rare, especially the scope of both wide and narrow.

Current uses: three uses were identified: 1) patients or health care providers searching for studies that are recruiting participants, 2) researchers conducting systematic reviews, 3) Advocacy groups, and various stakeholders accessing study information to display on websites tailored to particular audiences.

New uses: new uses and improvement suggestions include three themes: stronger search capability (search similar results, more powerful filtering/sorting), more user-friendly experience (walk through the steps for building a search query, indicate result relevance, display structured information of inclusion/exclusion criteria and make them more prominent.), providing guidance or information based on user types (customization of technical details, offer plain language summary).

Linkage of resources: as information on the web is connected, and most respondents requested for linkage to external medical resources such as PubMed citation or records of published journal articles containing study findings. Some proposed

linkage to the repository of studies' participant data and advocacy group websites.

Public Webinar

A webinar was hosted where attendants spoke out, and guest speakers shared their experiences and knowledge in the medical field. During the Webinar, live polls showed that most attendants voted for Google and another search engine (86%) over Doctor (80%) when it came to the question 'What sources do you use to find health information.' Moreover, the answers for 'Where do you look for information about the results of clinical trials' were clinicaltrials.gov summary results (71%), PubMed (57%), and Medical journal publication (63%).

There was a significant request for data visualization of study results and study design from attendants. Further, it was demonstrated by two guest speakers that organizing and visualizing the overview of all studies (Figure 38) that investigate the same intervention was critical, as well as binding study results. For instance, the Drug Facts Box (Schwartz & Woloshyn, 2013) introduced by one speaker, was designed to inform HCPs and patients about intervention benefits and risks, as well as help researchers better define research questions.



Figure 38. guest speaker demonstrating self-curated overview of all studies around one intervention https://clinicaltrials.gov/ct/about-site/modernization#modernization_rfi

3.3 User research (healthcare professionals)

This section strives to gain a deeper understanding of the user group and their current experiences with finding and accessing investigational treatment options (CT/EAP). The user research includes activities of literature reviews, surveying HCP, as well as expert interviews. The insights mapped out the current journey and illustrated two types of Search personas.

3.3.1 Empathizing healthcare professionals

HCP survey

A survey (appendix p. num) was spread among internal and external HCPs to learn HCPs' attitudes to CT/EAPs and their current solutions to resolve the problem of patients running out of treatment options. Ten (Figure 39) responded to the survey, including six external HCPs and four internal HCPs (medical team of myTomorrows).

(The survey was co-created with the UX designer, Melissa)



Figure 39. 10 HCPs including 4 from myTomorrow

Expert Interview

To supplement the aim of better empathizing HCPs with in-depth insights, semi-structured interviews (Figure 40) were conducted with experts in the medical field. A total of nine interviewees participated, scattering in 3 countries of the Netherlands, China, and Brazil.

One interviewee was a medical ethicist as the use of investigational treatment is one of her research areas, while the rest were all HCPs, and three were experienced specialists (HCP 4, 7, 8). All interviews were conducted remotely and followed the same interview protocol. For the experienced specialists, an additional section for collecting underpinnings of judgment on search results was

followed at the end of the interview (3.4.2 reasons for relevance feedback, p.66).



Figure 40. remote interview with HCP 8 in ZOOM

Interview protocol

- What would you do(know) when a patient ran out of treatment options?
- How to find investigational treatments (CT/EAP) now and any considerations?
- What would you expect from a tool/service that helps you find investigational treatments?
- How do you feel/think about giving explicit feedback?

Interviewees (9)

- Medical ethicist (NL)
- HCP 1, Physiotherapist, (BR)
- HCP 2, Physiotherapist, (BR)
- HCP 3, gynecologist, (BR)
- HCP 4, Psychiatrist (NL)
- HCP 5, Nurse, (CN)
- HCP 6, Epidemiologist, (CN)
- HCP 7, Dermatologist, (CN)
- HCP 8, Pediatric hematologist (NL)

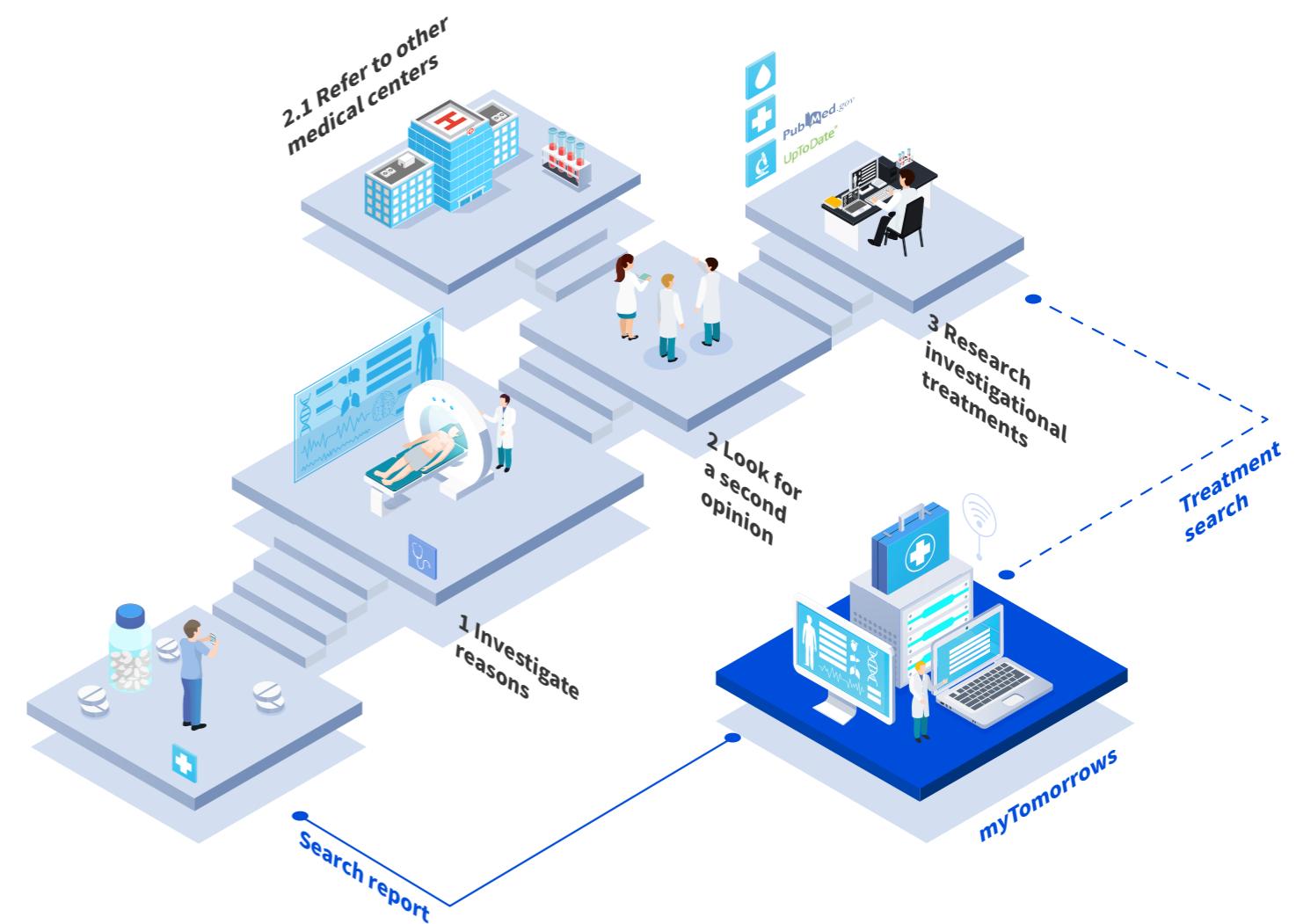


Figure 41. HCPs journey to CT/EAP (Drawn by author)

3.3.2 HCP journey to CT/EAP

Figure 41 illustrates a journey of actions HCPs take when a patient is not responding to or has run out of treatment options.

The first step for HCPs to take is to investigate the reasons that caused the patient not to respond, and they would try to adjust current treatment (e.g. dose). The next would be looking for a second opinion by discussing with peers or consulting other experienced HCPs. It is possible that the patient will be directed to another medical center.

However, turning to investigational treatments (CT/EAP) would be the very last step. In most cases, HCPs would turn to published articles from medical databases (e.g. PubMed) to research new

treatments because they include evidence or early data from patients. While some HCPs, especially those active in research and academia, may know all cutting-edge treatment developments in their specialized fields, thus, they will direct the patient to a CT/EAP that may benefit him/her.

Currently, myTomorrows involves itself in the journey from two ends. To HCPs, the treatment search engine comes into the journey as an approach to expand their scope of searching or mitigate the gap HCPs face from the current tools. On the other end, the search report produced by internal HCPs of myTomorrows plays an essential role in engaging HCPs through patients.

3.3.3 CT/EAP in HCPs' eyes

In the case of patients exhausting treatment options, there are a few identified facts and challenges HCPs face that may hold them back from prescribing investigational treatments.

HCPs may not be aware of CT/EAP as a treatment option

HCPs might not be aware of the possibility to prescribe drugs that are not yet approved for marketing. Only one HCP (out of six) surveyed directly mentioned that finding a clinical trial would be the next step to take when a patient is not responding or has run out of treatment options. Whereas, the rest would take actions to investigate the reasons or seek a second opinion. All HCPs knew CT, but it does not apply to EAP. The fundamental reason might be that the use of unapproved drugs outside of clinical trials is not taught in medical school or during residency. Instead, it is mostly self-taught (Irvine, 2016).

Survey HCP 1, ‘Will consider alternative diagnoses and do necessary investigations. Re-examine the patient in case (he) missed something. Check on history. If nothing is found, consult a colleague or seek a second opinion.’

Survey HCP 4, ‘I am not clear what exactly is meant by expanded access programme (EAP).’

Survey HCP 5, ‘In the case that I would not be the best-known expert on The Netherlands in his condition, I would first contact an expert in another center for advice.’

Most HCPs are open to CT/EAP but very cautious

Regardless of the fact of unawareness, HCPs are open to the possibility of prescribing investigational treatments to patients as the last option. All HCPs surveyed and interviewed said they might consider CT/EAP if it would potentially benefit patients, and patients are fully informed about benefits and risks. However, there are HCPs reluctant to prescribe EAP and wait for medical guidelines, even though there is promising data.

Survey HCP 2, ‘Yes (I will consider CT/EAP), as long as he is fully informed and no major risk to his condition.’

Interview HCP 4, ‘We normally stick to the guideline, and the farthest we go in clinical practice is off-label use of some drugs, but we know the efficacy pretty well.’

Interview Medical ethicist, ‘Even though there is promising data of a treatment, some doctors would wait for professional guidelines, which takes a long time after marketing authorization.’

Updates of CT/EAP are lagged

For HCPs who are aware of CT/EAPs, it could happen that they are not always up to date of what CT/EAPs are available because it requires a remarkable investment of time for systematic reading. Also, CT/EAPs are merely spread on a regular basis and within a limited scale.

Interview Medical ethicist, ‘Doctors in regional hospitals, they don't have much time to do any reading. While academic doctors would try to keep up with literature.’

Interview HCP 8, ‘I attend conferences, and we have a national working group. Of course I read journal articles. Sometimes, Twitter is a source for knowing new developments, and it's very fast.’

Unfamiliar with CT/EAP enrollment

Another challenge that may hold HCPs back is their unfamiliarity with the regulatory process, administrative burdens, finance and reimbursement, safety and liability concerns (Darrow et al. 2015). The findings are also supported by the results from the survey and interviews, as only a small fragment of HCPs (2) had prior experience with enrolling patients in CT/EAPs. Most perceived it challenging to enroll in a CT, especially with EAP.

Interview HCP 8, ‘We would send the patient to another center that runs clinical trial that could possibly benefit the patient, and I know these clinical trials because I am in the society in my field of specialization’

Survey HCP 5, ‘I would not say this is ‘challenging’. If the CT is not available in my center, I will refer the patient to another center. Things become challenging if there are only centers abroad that conduct the trial. I think depending on your personal contacts and experience, enrolling someone in an EAP can be challenging, but I have little to no experience with this.’

Uncertainty of trusting a third-party for finding and enrolling CT/EAPs

To most HCPs surveyed and interviewed, seeking help or assistance from a third-party for finding and enrolling would be acceptable. However, it remains to be skeptical unless HCPs are confident that such service providers are trustworthy.

Survey HCP 1, ‘I would consider using ‘a service’ when I know which one to use, when I know what this service does and when I have heard good stories about using this service.’

3.3.4 HCPs expectations on myT treatment search

A trust, unbiased and transparent platform

As research (Schwartz & Woloshin, 2013) has shown that published papers of drug efficacy include biased efficacy data. myTomorrows, as a commercial company, may give HCPs the impression that the information provided in the search is biased to pharmaceutical companies behind the scene. Therefore, HCPs would want a search to be trustworthy, unbiased, and transparent of how it works.

Interview Medical ethicist, ‘I think the first most important thing is probably trust. That's... It's really difficult to win outsiders. So how can you make sure that they trust that search engine? Probably by very carefully sourcing everything and explaining, doing a lot of explaining. So really showing where the data is derived from? Really make that complete, and it's relevant. And it's up to date.’

Interview HCP 4, ‘How does the company really provide services around the information from the search? I am afraid that it might be biased so that I will miss some other (treatment)...that might be better.’

A Complete CT/EAP database

As has been discovered from the future vision interviews (3.4 facts, p.56) that CT/EAP data is messy and scattered. HCPs expect the search to include all CT/EAP databases so that it brings much convenience in finding treatments or conducting

systematic reviews for research purposes.

Interview HCP 4, ‘I would certainly use it probably more for my research job, let's say just last month or two months ago, I had to write a mini-review for a grant proposal. Yeah, and then you need to check which research is already ongoing, and that involves going to at least three or four databases. So if you could do that with one overarching search engine, that would be a lot simpler.’

HCP characteristics



Altruistic

HCPs are willing to help their patients, even though they are busy in their daily routines.



Dependent

HCPs trust their colleagues and often look up a second opinion or collaborations (possibly from a different medical center).



Cautious

HCPs normally stick to medical guidelines when prescribing medicine, because of the uncertainty of using unapproved treatments. They need strong evidence before stepping out.



Critical

HCPs believe in scientific evidence (such as trial design and previous results of clinical trials).



All-time learner

HCPs need to keep learning and updating their knowledge in the medical field after medical school, for instance, systematic review.

3.3.5 HCP search personas



Susan (Active searcher)

Clinical oncologist

Age: 39

Current tools: PubMed, Uptodate, Medline Plus, etc

Story

It happens to Susan regularly that her patients are not responding well to the standard of care. She would try various approaches to help her patients and update her knowledge by reviewing published articles. Nevertheless, it is hard to keep it this way as she is also busy with patient visits and other daily routines, so it is a hassle for her.

She is aware of CT/EAP options, and she once referred a patient to another medical center for a clinical trial. However, she never tried expanded access programs for her patients because of unfamiliarity and other reasons.

Goal

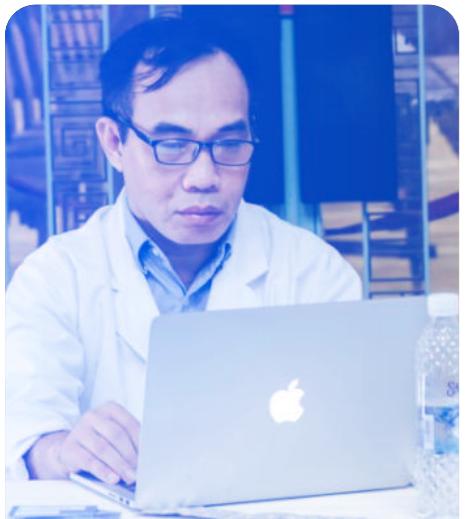
- To find alternatives for patients with unmet medical needs

Needs

- Search and review large amount of new treatment developments
- Evaluate possible treatment options for such patient(s)

Wants

- An easy way of finding alternative treatments
- Support for gaining access to CT/EAPs



Lee (Researcher)

Psychiatrist

Age: 42

Current tools: PubMed, clinicaltrials.gov or other registers, Social media (e.g. on Twitter), etc

Story

Lee is active both in clinical practice and academic research. In his field of expertise, it is common that patients run out of treatment options. Also, there are not a lot of new drug developments.

He is aware of all new developments as he is active in professional groups and attends conferences regularly, in which he will be updated systematically by peers and industry. As a researcher, he had experience with conducting clinical trials and writing academic papers that require systematic reviews.

Goal

- To gain a complete overview of research work related to his field of expertise

Needs

- Discover a complete set of new research works for research purpose for better defining the research question set up clinical trials
- Conduct systematic reviews

Wants

- A centralized database with high-quality research data
- An easy way of staying up to date of new research

3.4 Building blocks of feedback

It remains unknown what relevance feedback comprises. When, and where HCPs give such feedback during searches for treatment options. Thereby, studies were carried out to look into how HCPs conduct searches at present, to identify when and where HCPs make relevance judgments. Moreover, to discover what are the reasons for such judgments.

3.4.1 Treatment search flow

An observation session was conducted to learn the current search flow, behaviors, and identify where relevance judgments could be made while running searches for treatment options.

Approach

One internal HCP (treatment search coordinator) was recruited for this session because of the job function requiring searches and inaccessibility to external HCPs for real search observations.

The study subject was asked to record the desktop screen while searching for treatment options for a real patient (on clinicaltrials.gov). The screen record was the only material for analysis and was watched for three rounds. Each round served a different purpose, the first for immersion and primary understanding, the second for search flow generation, and the last for validating the visualized search flow.

Findings

Throughout the observation and analysis, it was learned that the treatment search flow aligns with information-seeking models (chapter 02, p.33). The search process (Figure 42) is repetitive and includes four phases from query to save and multiple iterative examinations of information. There are four main findings from this session as below:

- In order to have a higher recall, the search started from broad to narrow with more than one round of querying.
- Relevance judgments could happen when

scanning the search result list and examining the result content (blue dots in Figure 42). Irrelevant judgments are not always expressed and could happen merely in the head as some results were left 'ignored', while scanning through the search result list.

- It was needed to rely on web search engines (e.g. Duckduckgo) encountering information that is uncertain or beyond the searcher's knowledge scope.
- It was needed to refer back to a patient's condition and diagnosis when determining what trials are suitable for the patient. The nature of cognitive capacity may cause this.

Complete search flow in Appendices (p. 156-157)

Patient info (condition and diagnosis)

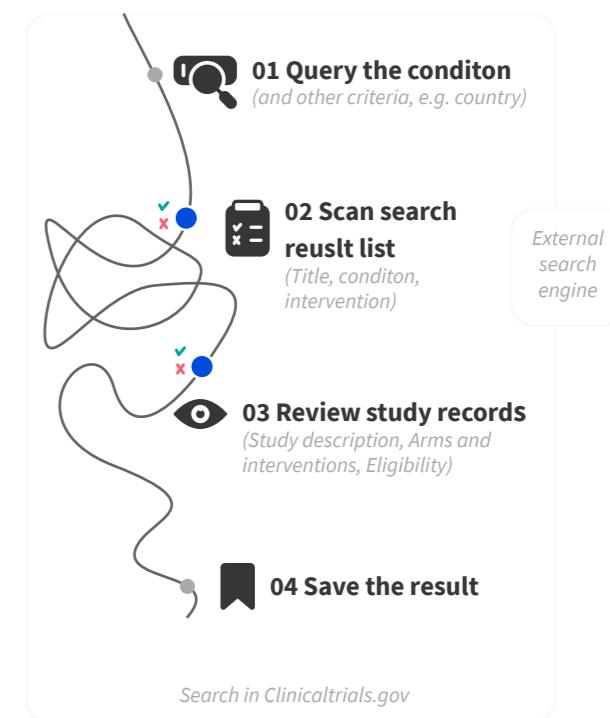


Figure 42. a simplified search flow overview (Drawn by author)

3.4.2 Reasons behind relevance feedback

An observation session was conducted to learn the current search flow, behaviors, and identify where relevance judgments could be made while running searches for treatment options.

Methods

Collecting reasons of HCPs' relevance feedback was conducted through role-playing sessions with three external HCPs (HCP 4, 7, 8 from expert interviews) and two internal ones (Figure 43). During the session, HCPs were asked to think aloud and prompted with the question 'Is this search result relevant or irrelevant, and please explain why?' while reviewing the list.

There are slight differences in the sessions with external and internal HCPs. Those with external HCPs were conducted in an imaginative scenario where a patient persona matching the HCP's field of profession was introduced, followed by presenting them with a list of six CT/EAP documents curated from clinicaltrials.gov. In comparison, the ones with internal HCPs (figure fixme) were done in an actual setting where they were searching for treatment options for a real patient with unmet medical needs.



Question:

Is this result relevant or irrelevant, and please explain why?

Figure 43. Role playing the AI while observing an internal HCP conduct searches (the rest were conducted remotely)

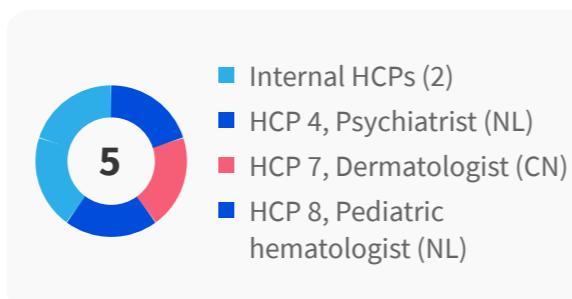


Figure 44. 5 participants including 2 internal HCPs

Feedback reason diagram

It was no surprise that there are three types of feedback: relevant, irrelevant, and not sure (table fixme). Various reasons were found for each type, but the reasons that build up irrelevant judgment were prominent and much more diverse than the rest.

Relevant: HCPs articulated reasons for a relevant judgment during the sessions. All external HCPs expressed a strong desire to know more about evidence, even though they thought it was promising. Whereas, internal HCPs will take actions on relevant ones (e.g., save).

Irrelevant: there are three types of situations where a searcher would make an irrelevant judgment: system failure (system failed to return correct results); insufficient expression (real information need is not put in the system, for instance; poor support of filtering or sorting), and flaw in the data (man-made problem with the original data).

Not sure: in addition to relevant and irrelevant judgment, unknown judgment was spotted because the information was beyond the knowledge scope of the searcher, or HCPs are hesitant or skeptical about a CT/EAP. The ultimate judgment that whether it will be relevant or irrelevant remains unknown, requiring further examination. Nevertheless, it proved that an undefinable judgment exists.

Figure fixme visualizes the feedback diagram with the query 'A', and it pinpoints where relevance judgment could be made in a search process.

+ Relevant	- Irrelevant	? Not sure
	System failure	
Relevant to query (Surrogate > Document > Information)	1) Irrelevant to the search query 2) Irrelevant to the filters applied Recruiting status such as age, gender location...	1) Beyond the searcher's knowledge scope (needing further research)
* Want to know more of its previous evidence	Insufficient expression 1) Irrelevant to the patient's condition : Diagnosis (Stage/Symptom), Eligibility criteria (Not specified) 2) Irrelevant to the patient's treatment need : Interventional type Efficacy of Intervention 3) Having access to a certain intervention ... Flaw in data 1) Outdated information (that fools the IR system) 2) Doubts on studies (Poorly designed study and purpose, lack of underpinnings or convincing data) ...	2) Hesitant or skeptical about trial benefits ...

Table 6. The reasons behind relevance judgment in three types

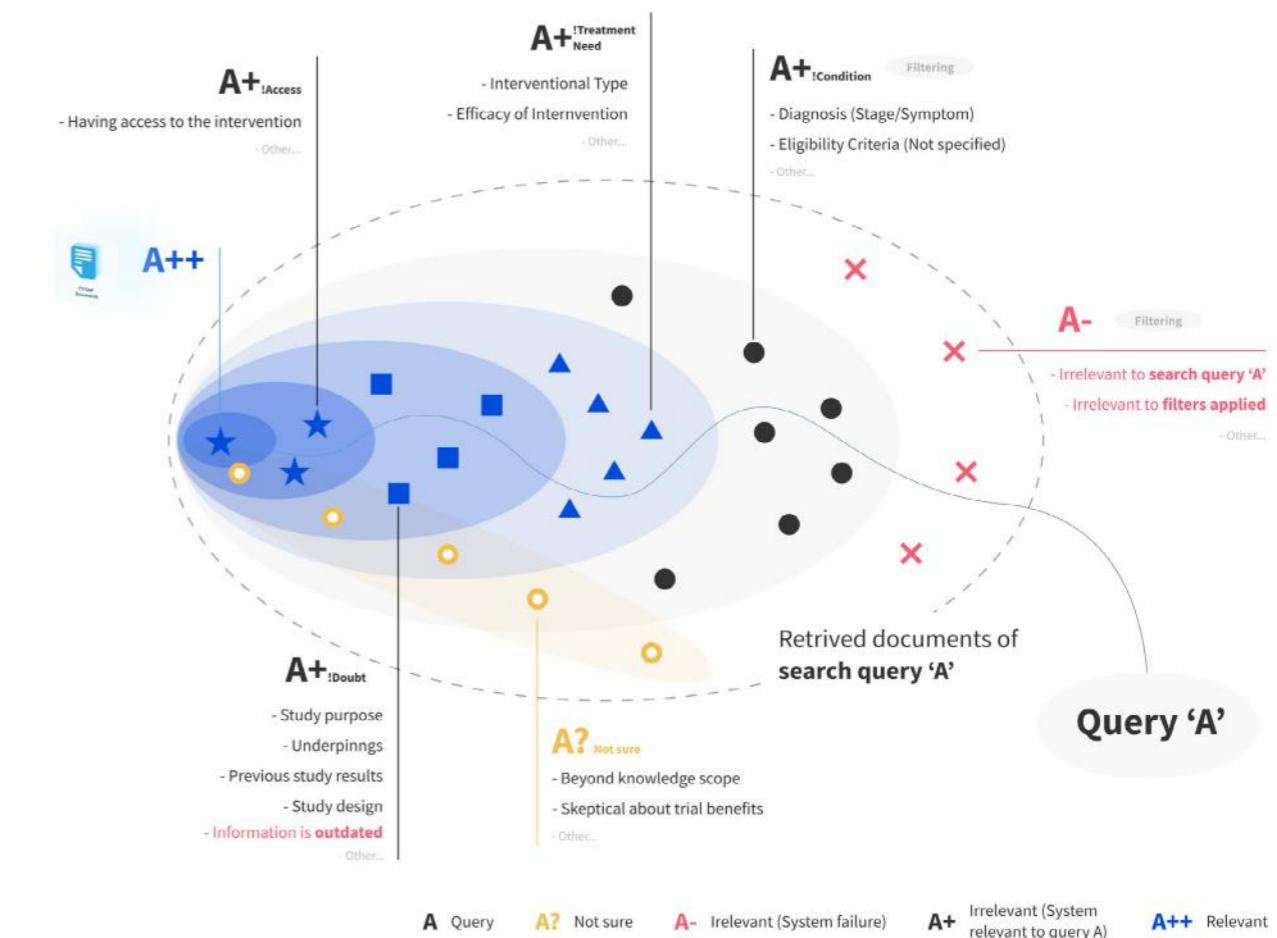


Figure 45. Visualized feedback classification diagram in the information seeking process

3.4.3 How HCPs perceive feedback provision

Some HCPs explicitly expressed their attitudes by sharing their own experiences during interviews. Together with the literature and the survey, there are two findings.

The ‘threshold’ of feedback provision

HCPs expressed that they felt neutral when giving feedback. Usually, they would not provide feedback unless it crosses the ‘threshold’ that the item is very relevant, or on the contrary, very irrelevant. It might be related to searchers’ emotional states influenced by the search results. Besides, HCPs felt that giving negative feedback is weird as they could ignore them, and for relevant ones, they can take necessary actions (e.g. save or download). The finding aligns with what researchers have discovered: giving negative feedback is perceived as unnatural (Druhv et al., 2019).

Interview HCP 3, ‘To be honest, I normally don’t really give feedback when it appears. But I do click only if I find something that is super useful or is what exactly I am looking for...or when it is a very bad article, that annoys me.’

Motivations for feedback provision is simple but hard to catch

According to the survey (Q10), HCPs chose ‘being informed about how providing feedback will benefit themselves and others’ as the top one in terms of being encouraged to provide feedback. The other two selected options were all about having controls on relevant (save) or (remove) irrelevant ones. Whereas, ‘receiving virtual reward’ and ‘being engaged in a competition’ was not selected.

Interview HCP 7, (translated from Chinese) ‘I don’t really think that I need to explicitly provide feedback, because I assume that if I click save or something it’s feedback right? For those I didn’t click or do anything, I think the system can regard this as irrelevant.’

3.5 Feedback collection in practice

User feedback is vital for digital experience improvement (Bajic & Lyons, 2011) in general. Paradigms of feedback collection are viable in most online products as suitable materials to learn from and discover underlying patterns. Fourteen examples of feedback collection in digital products were gathered and sorted depending on: the purpose (of feedback collection), triggering position, and collection methods.

3.5.1 Feedback purpose

Overall, the purpose of collecting feedback could be split into general and specific:

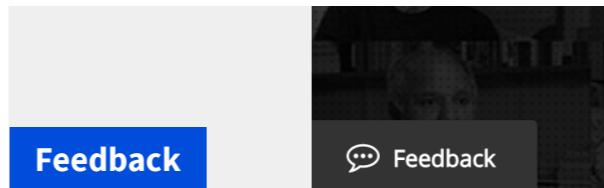
General feedback (4) is for gathering users’ overall experience, suggestions, or complaints with the product. Such feedback could be used for reflecting, improving product services, or identifying new business opportunities.

Specific feedback (10) refers to the feedback on a specific part of the content or feature (e.g. YouTube music collects likes/dislikes of a song). However, the goals for specific feedback varies distinctly from product to product. For example, Google Docs spell check collects feedback on its correctness of suggestions, while Microsoft Teams collects the quality of video calls. To this project, the collection relevance feedback is seen as specific feedback.

3.5.2 Triggering position

Triggering position means where the UI for feedback collection appears. Three types of triggering positions were identified: Fixed, Contextual, and Attached. The positioning of feedback collection UI is primarily related to feedback collection purposes, for instance, Fixed appears only for general feedback collection.

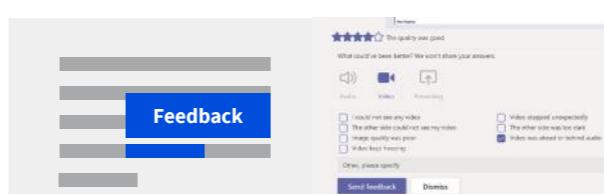
Fixed (4) refers to those feedback UI components fixed in an interface, commonly appearing in the corners (e.g. ACM digital library), and could be found only for general purposes.



Attached (6) stands for those feedback UI components that are attached to an individual content element. For instance, Google search (featured snippet) has a feedback button on the bottom right of the snippet.



Contextual (4) refers to those feedback UI components that appear to users only on certain conditions such as the completion of a task (e.g. Teams shows a contextual feedback request in some chat after a video call)



3.5.3 Collection methods

There are five specific methods for gathering user feedback in these examples, and usually, a feedback request combines different methods.

Binary judgment (5) (e.g. likes/dislikes)



Scale (5) (e.g., rating)



Predefined options (6) (e.g., reasons for video call quality)

- Option 1
- Option 2
- Option n

Free textbox (10)

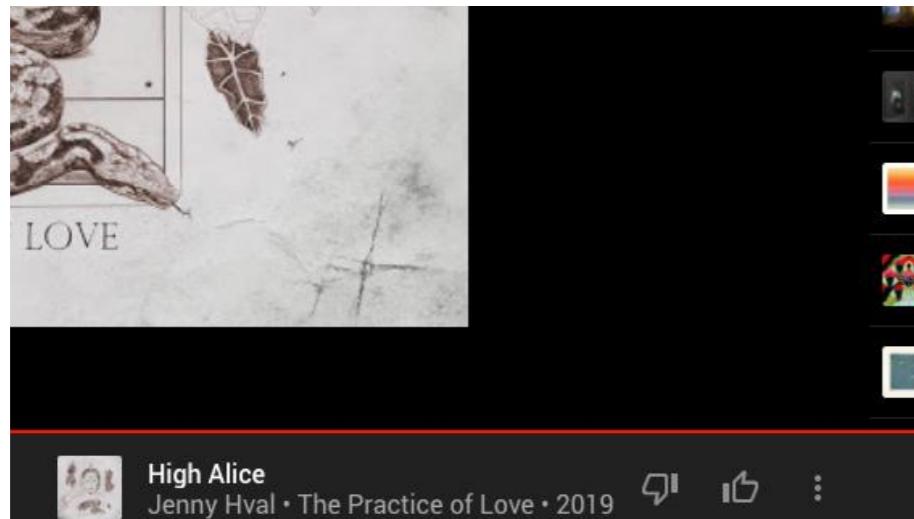
Please tell us why
(optional)

Pointing out (4) represents the feature allowing users to select a specific area or a part of the content on which users give feedback. (e.g. Users to capture a screenshot of a specific area in Google Forms)



Free text is essential for feedback as ten products incorporate it, and it is a must-have to all (4) those aiming for general feedback, and Scale comes next (3/4) for general feedback collection. For specific feedback, binary judgment and predefined options are vital (Table 6 could be translated into predefined options for this project, p.67). However, Pointing out appears only in four products, but it applies to both general and specific feedback collection cases. It could be because pointing out could help add a detail level to user feedback.

Youtube music



Purpose	General (4)	Specific (10)		
Triggering position	Fixed (3)	Contextual (1)	Attached (6)	Contextual (4)
ACM digital library, Hotjar, Google forms	Figma	Amazon sponsored results, Google photo, Google help, Google search (knowledge graph), Google search (featured snippet), YouTube music, YouTube	Google meet, Teams (video call), Google docs (spell check)	

Google help

These ads are based on the content of the videos you've watched whether you're signed in or not.

Manage privacy settings for ads
You can control the ads that you see based on your Google Account [Ad Settings](#). You can also [view](#), [delete](#), or [pause](#) your YouTube watch history.

Turn off ads
If you want to turn off ads on YouTube, check out our [paid memberships](#) for an ad-free experience.

Fill in lead forms
When you see and open a lead form in a video campaign on YouTube, some fields in the lead form are [pre-filled](#) when you're signed in to your YouTube account.
[Give feedback about this article](#)

Was this helpful?

Collection methods	Binary judgment (5)	Scale (5)	Predefined options (6)	Free textbox (10)	Pointing out (4)
	✓				
		✓			
			✓		
				✓	
					✓

Table 7. Fourteen feedback collection examples sorted in categories.

Youtube



Table 7 shows an overview of sorted examples (14) by three categories, as discussed earlier. Those under the category Specific serve as useful references for ideation on relevance feedback collection interactions.

For specific feedback collection, the Attached triggering position is more common than Contextual. However, it does not imply that the former is better than the latter, which could be a result of differences in user flows or behaviors within these products. Moreover, Fixed may be more suitable for general feedback collection,

and it requires extra steps (e.g. select or capture a screenshot) to link feedback to a specific part of the content. Thus, the ideation phase will avoid using a fixed position for relevance feedback.

All five collection methods, except scale, will be considered in the ideation phase, for the reason that the previous section has identified relevance is not a scale of preference, but more of a binary judgment.

* Appendix (p.164-167) contains all examples and a comparison form of all fourteen products.

Key takeaways chapter 03

3.1 Product analysis

The knowledge graph of myTomorrows search plays an essential role in the search, and it connects two data sources. One data source for building the knowledge graph and the other for CT/EAP source documents. There is only one CT/EAP data source incorporated in myTomorrows search, but seventeen other recognized sources are excluded. There is no ranking of search results in myTomorrows search engine result page, but it is not communicated to users, which may inherently lead to a biased results perception.

The information architecture, together with the heuristic evaluation results, has seen a series of UX issues that need to be addressed in the future concept design, and the focus should be on improving the flexibility and efficiency of use.

3.2 Future visions on search

The future visions on the treatment search from myTomorrows perspective are categorized into two themes of near-future vision and far future vision. Those in near-future visions, including eleven points: Complete source data, Quality & rich information, Clear actionability, Control & feedback on results, Result management, Search report manipulation, Insights & trends, Discussion & co-decision between HCPs, More user-friendly, Patient search as an option, and Lower search barrier. They can be incorporated in the future concept design in this project. The far future visions, in comparison, require a longer time frame and are beyond the scope of the project.

Besides, through the longitudinal study on Clinicaltrials.gov (Modernization), which shares a similar group of end-users, it is learned that their user requests are mostly on unifying the CT/EAP document standard and a more user-friendly system. The latter is somewhat in line with the future visions of myTomorrows.

3.3 User research (healthcare professionals)

Through an online questionnaire with six external HCPs and four internal HCPs, together with nine in-depth interviews with external HCPs, it is learned that for most HCPs, choosing to turn to investigational treatment options (CT/EAPs) are rare and is often the last step to take in clinical practices. Despite this, it is acceptable but cautious for HCPs to recommend their patients to participate in CT/EAPs. However, some HCPs are unaware of such options, and most are not familiar with enrollment procedures. HCPs, as a highly educated group, share common characteristics such as Altruistic, Dependent, Cautious, Critical, All-time learner. Those characteristics could help inspire the design in the following stages.

HCPs expect a search-based platform (myTomorrows search) to be a transparent, unbiased, trustful platform, and it could provide a complete CT/EAP data. The two fundamental expectations reflect on the search Personas as Active searcher and Researcher. The former wishes to help the patients in need without bringing potential issues or offering false hope; in contrast, the latter aims to have a full picture of the cutting-edge research works.

3.4 Building blocks of feedback

By observation, it sees two places where end-users make relevance judgments. One is on the search engine result page where users prescreen search results, and the other is on the search result content page where users read through the content for a deeper level of understanding to make relevance judgment. These are the two critical points from which relevance feedback to be collected.

Results from the role-playing sessions discover the determining factors of relevance judgment that users' relevance judgment could be classified

that users' relevance judgment could be classified as relevant, irrelevant, and not sure. There are three main reasons identified irrelevant judgment: system failure, insufficient expression (of the information need), and flaw in the source data. For the not sure judgment, it is the result of inadequate knowledge users have to make judgments or the hesitance they might have. The reasons that determine relevance judgment will be used for building the pre-defined options for the feedback collector.

3.5 Feedback collection in practice

In practice, explicit user feedback collection can be generally classified by three parameters: the purpose, triggering position, and collections methods. Typically, the purpose of collecting user feedback is either for general purposes (e.g. evaluation user satisfaction) or specific purposes (e.g. machine recommendation).

The purpose determines how feedback collectors position and what type of methods used to collect user feedback. For specific purposes (as in line with this project), attached and contextual positions are more common. In comparison, the fixed position can only be found for those with general purposes. Moreover, for specific purposes, all methods (binary judgment, scale, pre-defined options, free textbox, and pointing out) are incorporated.

04

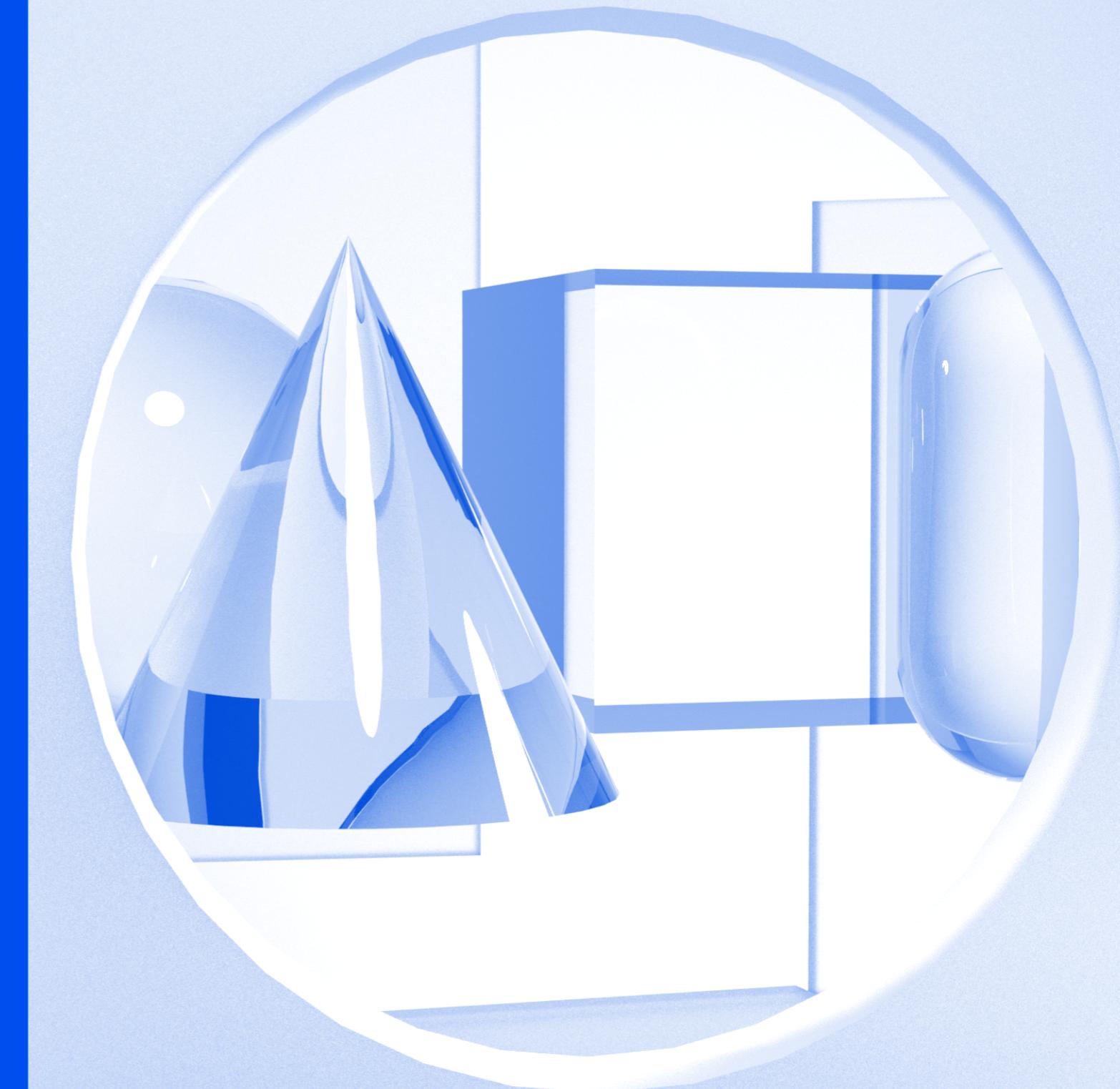
research synthesis, envisioning the future



Chapter four synthesizes the findings and insights from the theoretical and exploratory research done in the previous two chapters, aiming at consolidating a strategic solution to tackle the challenges of this project. Under the design goal, four qualities in design were mapped out and translated into four design criteria, accompanied by specific lists of requirements.

Chapter overview

- 4.1 Design goal
- 4.2 Design criteria and requirements



4.1 Design goal

Design a future myTomorrows treatment search interface concept to engage healthcare professionals to provide explicit relevance feedback.

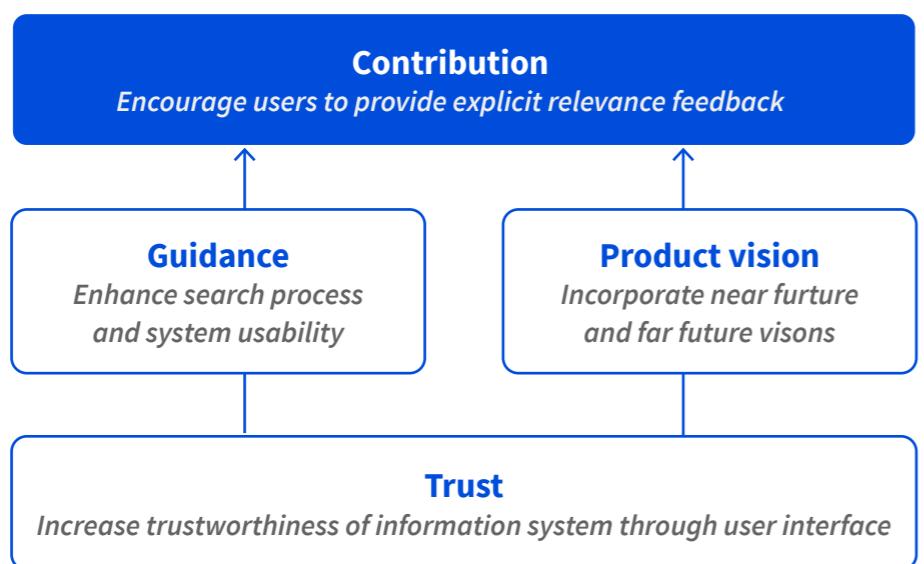


Figure 46. Four qualities towards the design goal in this project. (Drawn by author)

4.1.1 Design goal breakdown

The design goal is a synthesized statement striving to resolve the two identified challenges of the project assignment (the unmet urge for UX improvement and the lack of user feedback data for system optimization, p.22).

Previous research has found that to lead users into a particular behavior (in this project, providing explicit relevance feedback on search results), the information system needs to be highly useful and usable (behavior model, p.40). More importantly, users should be able to sense the credibility of an information system, or the perceived trust (engagement theory, p.44).

Therefore, to reach the ultimate goal of user contribution (users provide explicit relevance feedback), three other qualities need to be addressed: trust, guidance, and product vision.

Figure 46 illustrates an interrelated relationship between them. Trust is supposed to be the fundamental quality to achieve, without trust, Guidance and Product vision would be trivial, not to mention Contribution. Guidance and Product vision are the key to users that it supports users to reach their goals, impact behaviors, and creates different paths of leading them to provide relevance feedback

Trust

It refers to the quality of users' perceived trust or credibility in an information system. Many personal and impersonal factors influence the construct of trust (p.44), but to the scope of this project, it will be solely focused on the dimension of system trust (Mcknight, 1998). More specifically, increasing the transparency and interpretability of the information system and improving the visual, interaction, and information of the user interface.

Guidance

It stands for the quality of how well an information system performs to support users to achieve their intended goal of seeking relevant information conveniently. As is the core of a search system, the information seeking process is simple but complex (p.30) as of its dynamic, repetitive, and iterative nature. It could be enhanced by assisting users in formulating information needs, expressing needs through querying, evaluate search results, and act on relevant information. Besides, usability is another key to enhance Guidance. The current myTomorrows search sees abundant UX issues (heuristic evaluation, p.52) and opportunities for improvement (p.28).

Product vision

It means the quality of users' perceived usefulness and functionality of an information system, mainly achieved by product definition of the treatment Search. In other words, future visions. It is closely tied to Guidance, and the change of future vision

leads to different or new guided user behaviors in Search since user behavior is influenced by the functionality of the web browser software (Choo, Detlor, & Turnbull, 2000).

Through future vision interviews with myTomorrows' employees (p.54), the longitudinal study on clinicaltrials.gov modernization (p.57), and user research on healthcare professionals (p.58), a variety of visions and expectations were gathered.

Contribution

Alternatively, relevance feedback provision. It stands for the quality of engaging users to make contributions by providing explicit relevance feedback on search results. As is the ultimate objective and the determining quality of the design goal, Contribution is built on top of the other three qualities. Without achieving the other three, Contribution can not be maximized or even achieved.

Previous research on relevance (p.36) and information seeking process (p.30) helps identify when and where relevance judgment comes from, and the reasons for relevance judgments (p.65) answer the question of what. On the other hand, research on motivation-related theories (p.39) and explicit feedback collection in practice (p.66) shed light on why users would take actions to provide feedback and the design of such interactions.

However, it remains uncertain how and what motivators could better engage users to provide relevance feedback. The following chapters (chapter 5, 6) will be focusing on answering the two questions by designing and speed dating ideas with users.

4.2 Design criteria and requirements

On top of the previous subsection, which introduces the relationship of the four qualities that the design goal should strive to reach, this subsection aims at consolidating each quality by synthesizing the findings and insights from previous theoretical and exploratory research. As a result, it forms four design criteria with specific requirements.

O1

The design should increase users' perceived trust on myTomorrows search system through its interface.

The first design criteria is a translation of the quality Trust, as mentioned earlier. It should aim to achieve the list of requirements (table fixme) belonging to the two levels of information system and user interface. Both will be reflected in the user interface, but information system focusing on users' perceived system capability and trustworthiness. In comparison, user interface is a mixture of information representation and overall impression.

	Requirements	Reference
Info System	#1.1 Explain myTomorrows search intention and how it works	(NN group, 2016, p.44) (Fogg, 2001, p.40)
	#1.2 Display (CT/EAP) source databases and allow free selection of them	(Gil, Y., & Artz, D. 2006, p.44) HCP expectations, p.61
	#1.3 Provide explanation of the ordering mechanism of the search results	(Morville 2010, p.29) Product analysis, p.50
	#1.4 Indicate system search performance comparisons to (CT/EAP) source databases	HCP expectations, p.61
User Interface	#1.5 Display and provide the paths to (CT/EAP) source databases	(Fogg, 2001, p.40) HCP expectations, p.61
	#1.6 Explain the size of search results	(Gil, Y., & Artz, D. 2006, p.44)
	#1.7 Highlight information recency	(Gil, Y., & Artz, D. 2006, p.44)
	#1.8 Provide consistent and professional (health-related) imagery	(Gil, Y., & Artz, D. 2006, p.44) (NN group, 2016, p.44) Heuristic evaluation, p.52

Table 8. List of design requirements aimed at increasing perceived trust (Note that the ordering does not represent the priority)

O2

The design should smoothen users' search process and enhance guidance to relevant information with lower cost.

	Requirements	Reference
Need formulation	#2.1 Assist users to specify (narrow down or expand) their information needs	(Hearst, 2009, p.28) (Morville, 2010, p.28)
	#2.2 Enable advanced search	(Morville, 2010, p.28)
Query	#2.3 Display query history	(Hearst, 2009, p.28)
	#2.4 Suggest related queries and enable query autocomplete	(Hearst, 2009, p.28) (Morville, 2010, p.28)
Evaluate	#2.5 Indicate reviewed and unreviewed search results	(Hearst, 2009, p.28)
	#2.6 Display and provide the paths to (CT/EAP) source databases	(Hearst, 2009, p.28) (Fogg, 2001, p.40) HCP expectations, p.61
Act	#2.7 Allow varied actions on search results (e.g. save, download, share)	(Hearst, 2009, p.28) (Morville, 2010, p.28)
	#2.8 Provide flexible access to contact myTomorrows	(Hearst, 2009, p.28) Future vision, p.54
Info scent	#2.9 Allow customized display of information in search result page	(Pirolli, P., & Card, S., 1999, p.34) Treatment search flow, p.63
	#2.10 Offer control to users on every actions	
Usability	#2.11 Respond to controls with immediate and accurate feedback	Heuristic evaluation, p.52 Clinicaltrial.gov modernization, p.57
	#2.12 Shorten the path to review information	
	#2.13 Allow customized search queries	
	#2.14 Improve the information readability and reviewing efficiency	

Table 9. List of design requirements aimed at smoothening search process and enhancing guidance (Note that the ordering does not represent the priority)

This is a representation of Guidance. The listed requirements are based on the search process (need formulation, query, evaluate, and act), search strategy, and the identified UX issues from heuristic evaluation and users' urge.

03

The design should incorporate feasible product visions of myTomorrows and expectations of healthcare professionals for the near future.

This design criteria is equivalent to Product vision. The requirements mostly comes from the future vision interviews and the longitudinal study of clinicaltrial.gov modernization. In general, they are divided as near future and far future in accordance to myTomorrows product roadmap. However, this project will solely be focused on incorporating the requirements in the category of near future due to the limited time and project scope.

Near future

Requirements	Reference
#3.1 Include complete (CT/EAP) source databases	<i>Future vision, p.54</i> <i>CT/EAP in HCPs' eyes, p.60</i>
#3.2 Offer educational information of CT/EAP and access	<i>Future vision, p.54</i> <i>CT/EAP in HCPs' eyes, p.60</i>
#3.3 Curate or linkage to published CT/EAP efficacy data	<i>Future vision, p.54</i> <i>Clinicaltrial.gov modernization, p.57</i>
#3.4 Access to and manipulate interactive search reports	
#3.5 Support collaborative decision making between HCPs	
#3.6 Support search results management	<i>Future vision, p.54</i>
#3.7 Offer a simplified search for patient users	
#3.8 Enable user authentication	
#3.9 Notify updates of interested disease area	
#3.9 Open the access to patient data input to run search directly	<i>Future vision, p.56</i>
#3.10 Help match participants for clinical studies (two-way matching)	

Table 10. List of design requirements of product visions (Note that the ordering does not represent the priority)

04

The design should encourage users to provide explicit relevance feedback without the use of material incentives

Ultimately, the last criteria (The same as Contribution) is the most important one. The requirements listed below blend findings and insights from theories and in-depth research with end-users. This list may not be comprehensive as mentioned in the previous section, but it serves the foundation for further research by designing.

Reference

Requirements	Reference
#4.1 Align the value of providing feedback to motives	<i>(O'Brien & Toms, 2008, p.39)</i> <i>(Attfield et al., 2011, p.39)</i> <i>(Fogg, 2007, p.40)</i> <i>Theories in motivation p.41</i>
#4.2 Distinguish the trustworthiness of the feedback from different users	<i>Future vision, p.54</i>
#4.3 Communicate the efforts needed for providing feedback	<i>(O'Brien & Toms, 2008, p.39)</i> <i>(Attfield et al., 2011, p.43)</i> <i>(Fogg, 2007, p.40)</i> <i>(Thaler & Sunstein, p.40)</i>
#4.4 Respond immediately when feedback is provided	<i>(O'Brien & Toms, 2008, p.39)</i> <i>(Attfield et al., 2011, p.39)</i>
#4.5 Allow modifying a given feedback	<i>(Mizzaro, 1998, p.36)</i> <i>How HCPs perceive feedback provision, p.66</i> <i>Reasons behind relevance feedback, p.65</i>
#4.6 Offer assistance when negative feedback is given	<i>(O'Brien & Toms, 2008, p.39)</i> <i>(Attfield et al., 2011, p.39)</i>

Table 11. List of design requirements of guiding the design of feedback collection (Note that the ordering does not represent the priority)

05

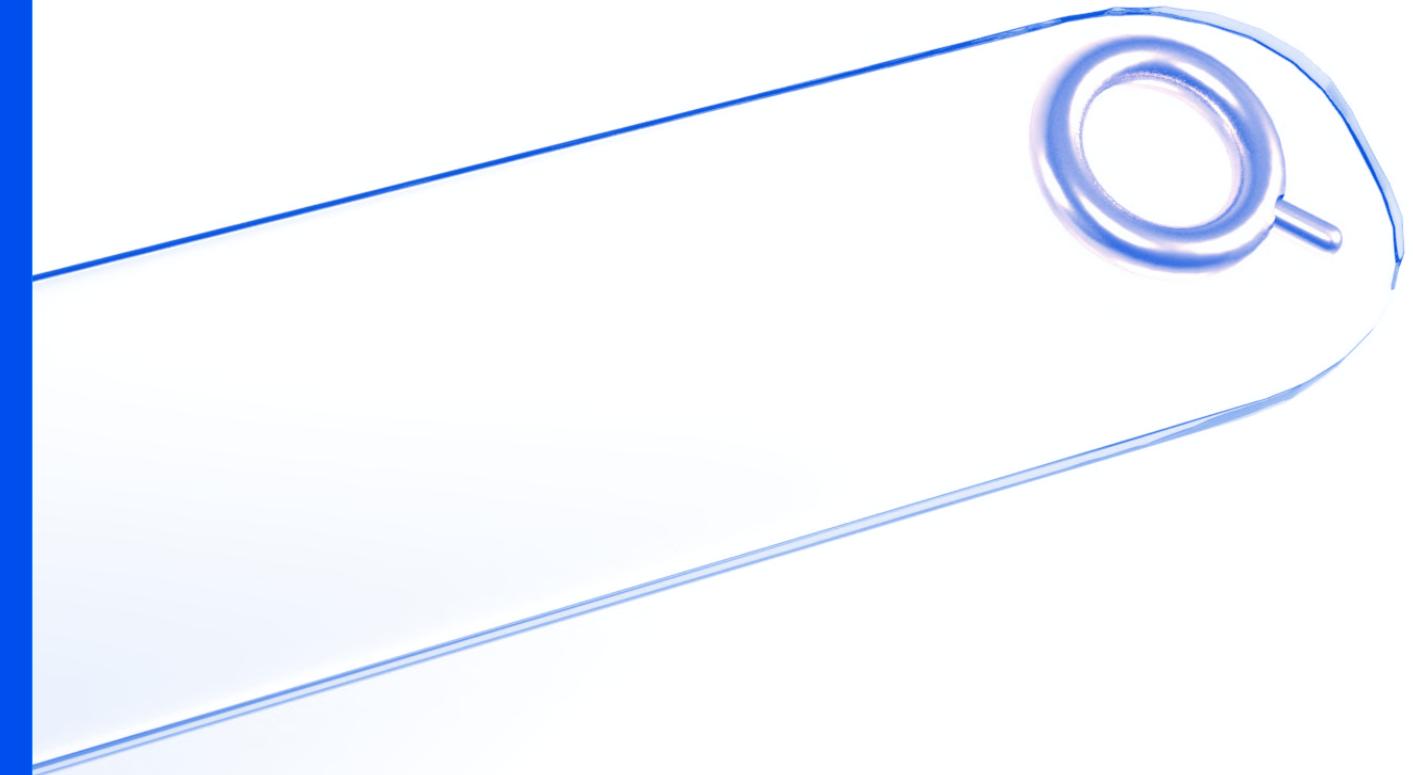
Future search conceptualization



This chapter aims to conceptualize and design the embodiment of myTomorrows treatment Search that collects explicit relevance feedback. It went through an iterative process and started with a creative session with peers to gather interaction ideas to supplement the design requirements (chapter 4). A baseline search was constructed as the foundation for feedback ideas realization by first modifying the information architecture, and creating the main user flows. Nine feedback ideas came along within the user flows. All nine ideas, together with the baseline search, were speed dated and tested for usability with peers and internal HCPs of myTomorrows to lay the basis for feedback conceptualization. Eventually, three feedback collection concepts embedded with different motivators were synthesized in the same baseline search.

Chapter overview

- 5.1 Creative session for design ideation
- 5.2 Baseline search construction
- 5.3 Feedback ideas generation
- 5.4 Usability testing and ideas speed date
- 5.5 Feedback enhanced search conceptualization



5.1 Creative session for design ideation

One creative session was carried out in order to generate initial interaction ideas that could potentially achieve the intended effects (Trust/Guidance/Contribution) in the treatment Search. A variety of ideas were collected in the end, inspiring the design of the Search interface.

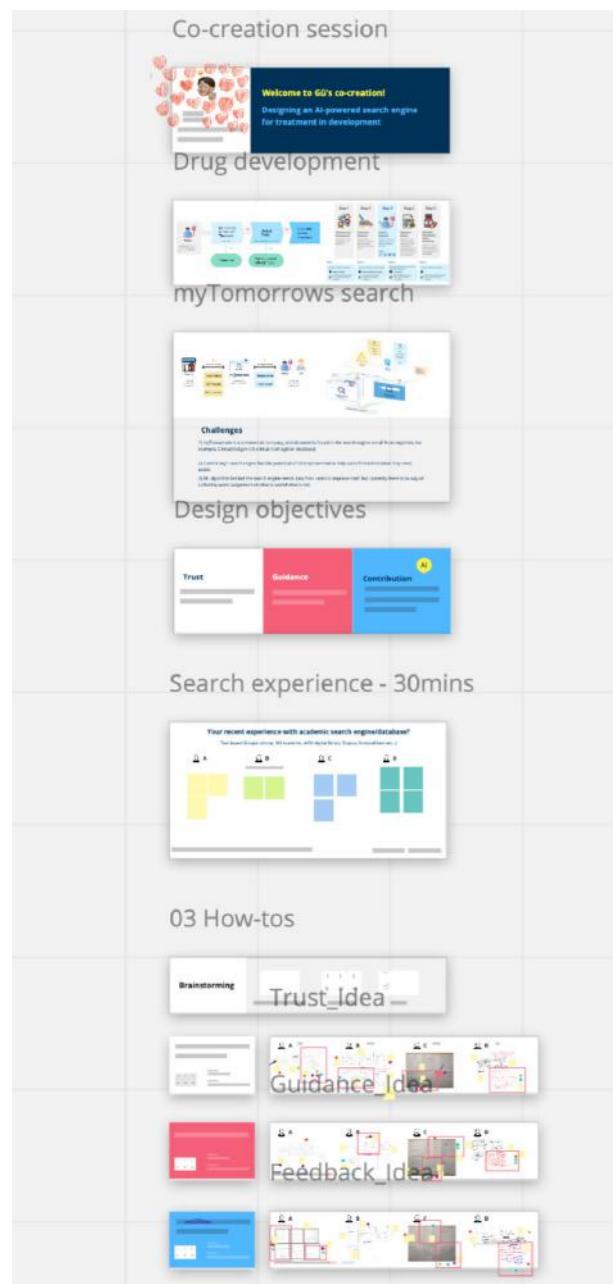


Figure 47. Screenshot of co-creation session on Miro

5.1.1 Approach

Four peer students specialized in design were recruited because of convenience and the unavailability of healthcare professionals.

The session (Figure 47) was hosted in Miro (an online collaborative tool) and ran through four sections (Figure 48) in approximately one and a half hours. It started with a context briefing of myTomorrows and the need to search for CT/EAPs, followed by a discussion section to help participants better immerse the Search context. During the discussion section, participants were asked to share their own experience with web search, especially with academic databases (e.g. Google scholar), which resembles how healthcare professionals use search engines in the medical context.

How-tos was employed to brainstorm ideas on the three qualities of Trust, Guidance, and Contribution in the search (Product vision was excluded due to their lack of medical knowledge). For each quality, three subsections were hosted, and each comprises two phases. The first phase focused on generating initial three ideas within one minute, while the second emphasized the iteration of previous ideas in a more extended period time of three minutes.

After the completion of How-tos was the reflection section, participants explained their ideas, discussed freely, left feedback, and voted their favorite concepts with three votes for each subsection.



Figure 48. Procedure of the creative session

5.1.2 Creative session results

The results of the creative session comprised two parts. One part is the experiences participants had over the three qualities of Trust, guidance, and contribution. The other is the interaction ideas around the three qualities mentioned above.

Experienced trust

Most (3/4) participants perceive trust in search engines on the information level. Its recency, the use of the information were two aspects heavily emphasized.

P1, "Date, year is really important because old ones can not really be used. Also (the number of) citations and reviews (in other cases) are very important."

P3, "the date of document is what I care much about"

However, one participant (P2, who had professional working experience with AI systems)

expressed his experience of trust in the system level. He addressed the explainability and control over IR systems would be essential factors for convincing users to trust.

P2, "I would feel more trust in the system if it allows me to edit or give more control over what I see...Also, if the AI system would be able to explain how the decision (of listing search results or recommendations) are made."

Experienced guidance

All participants expressed that their search process was dynamic, exploratory, and repetitive, which revealed a strong connection to the results from literature research. (search as part of information seeking, chapter 02).

P1, "I normally start with Google scholar and then depending on documents to explore deeper, or just use other sources."

P2, "I think running searches isn't so complex that requires a lot of guidance in the system, even though I don't always find what I need. Maybe filters help a lot."

Participants didn't expect much guidance in search systems helping search better, and would reflect on their own competence, this might be the result of poorly supported guidance in available search engines. But the need does exist as one participant addressed the barrier while using advanced search.

P4, "I think my knowledge base really affects my searches."

P3, "I use advanced search such as Boolean search because I have ever taken a course about that, but I don't always remember exactly how to do it."

Experienced contribution

The notion and experience of contribution by providing explicit relevance feedback in search engines was a little novel for all participants as none of them had prior experience giving feedback in such a context.

P4, "It sounds a bit odd that why should I provide feedback on my search results? I never encountered this before while searching."

How-tos - Trust

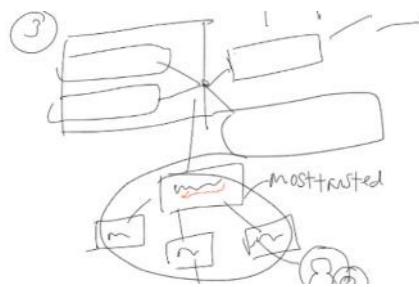
13 ideas were collected, and all concepts strongly relate to the previous findings that trust falls on the information, as well as the system capability. Ideas were grouped into 3 categories.

Displaying collective relevance (7 ideas): judgment from others matter since these ideas include showing popularity, citation, likes, ratings, comments, and credible individuals/entities consuming information.

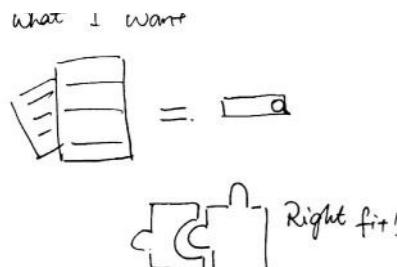
Increasing system explanation (4 ideas): knowing how search systems work sets a proper expectation. Ideas were about explaining how search systems function to match results to keywords and the boundaries of system capability.

Enhancing search functionality (2 ideas): helping participants better search would increase trust. Ideas comprise visual search and interactive query expansion/suggestion.

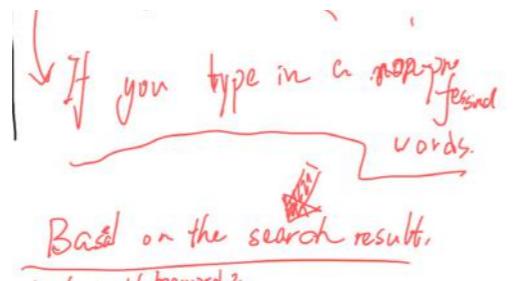
Top 3 ideas



(4 votes, Displaying collective relevance) Visualized overview tree of search results in combination with collective relevance. (likes, citation, etc.)



(3 votes, Increasing system explanation) Providing explanation of search result matching.



(2 votes, Enhancing search functionality) Extraction of search keywords from search results.

How-tos - Guidance

15 interaction ideas were generated, showing a wide diversity in these concepts that fall into 5 categories.

Interactive result matching (5 ideas): as the essential goal of running searches is to find relevant information, ideas in this category focus on using retrieved search results to find better or similar results, for instance, a Tinder-like swiping to generate better results, adding to a reading list based on which recommendations appear.

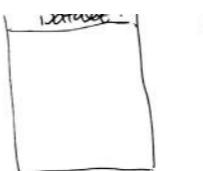
Recommending search queries (4 ideas): participants generated various ideas of machine recommendation of search queries. For example, displaying a visual graph with connections to other related keywords and showing real-time suggestions while typing.

Linking external resources(2 ideas): two ideas describe that external resources might be linked to search results to provide a more comprehensive overview and resolve situations where users could not capture or understand retrieved information.

Improving reading efficiency (2 ideas): it was addressed that when reading long text-based results. For instance, participants preferred to have valuable content highlighted (e.g., Medium) to improve reading efficiency.

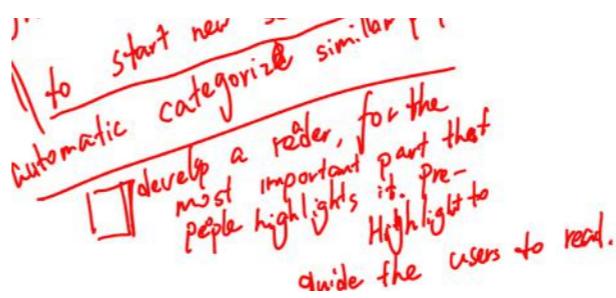
Personalizing search process(2 ideas): users have different needs for search. The ideas in this category emphasize that search should enable customized search procedures for varied purposes, such as allowing the creation of customized search templates.

Top 3 ideas



Develop My Personal Dataset

(4 votes, Personalizing search process) Customizable templates of how to conduct searches.



(3 votes, Improving reading efficiency) Show a 'heat map' within a document in an interactive way so searchers can find key information quicker. (with visual elements)



Provide relevant paper
based on your search history.

(2 votes, Recommending search queries) Show recommendations (multimedia content) based on search history.

How-tos - Contribution

Though participants didn't have experience of giving explicit feedback in search, 15 concrete ideas were generated belonging to 3 categories.

Feedback as a function (7 ideas): participants would want to see values when giving feedback, it could either help them expand search queries or increase the utility of the search engine. For instance, five ideas were about sorting search results into different baskets with pre-defined reasons, helping better manage results.

A simple question (5 ideas): participants were familiar with a question appearing on the screen

asking for feedback outside the search context. Such ideas were brought here in different stages in the search, such as asking the relevance question at the end of reviewing.

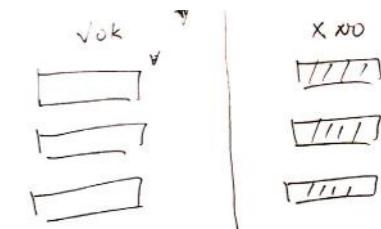
Reciprocity feedback (3 ideas): participants perceived giving feedback as additional and unnecessary when interacting with search engines, but it was more acceptable to provide feedback when taking actions on search results(s). For example, when users download or save a search result, they will be asked to provide the feedback by selecting reasons, adding a tag, or leaving a comment.

Top 3 ideas

Thank you for
Downloading
To help contribute
the knowledge --

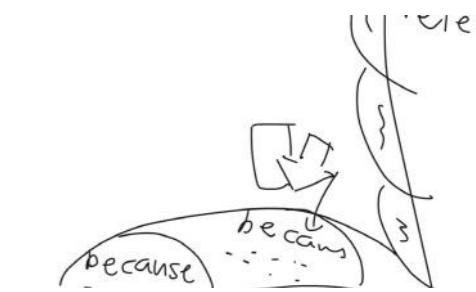


(4 votes, Affiliated feedback) Combining feedback request when users take actions (download, save, etc) on search results.



I also asked the user

(3 votes, Feedback as a function) Record the search process and in the end, ask search (from the search overview) what is relevant, what is not and why.



(2 votes, Feedback as a function) Drag and drop to customizable baskets, so results can be determined what to do with later.

5.2 Baseline search construction

This section aims to create a baseline search interface in terms of Trust, Guidance, and Product vision, serving as the foundation for the realization of Contribution. It starts with the reconstruction of information architecture, based on which is the creation of the primary user flows.

5.2.1 Information architecture reconstruction

The information architecture reconstruction adapts from the original one (p.51) to one that meets the design criteria (chapter 4). Figure 50 shows the revised information architecture. Branches colored blue remain unchanged as the original one, and most of the changes derive from the third design criteria of product vision.

Main changes

1. **Displaying myTomorrows LOGO and footer information alongside Treatment search LOGO:** keeping HCPs in the flow of search while being connected to the main website that includes detailed information about myTomorrows services. (#2.12, p.77)
2. **Search report:** bringing the current search report (PDF) online to enhance HCP engagement and increase the exposure of myTomorrows search. (#3.4, p.78)
3. **Subscription:** allowing users to stay updated (with alert) in the fields of interest. (#3.9, p.78)
4. **Account:** enabling user authentication to support interactive search reports (generation) and information protection. (#3.8, p.78)
5. **myFolders:** empowering users to manage saved search results, especially for internal HCPs, to reduce repetitive workload. (#3.6, p.78)
6. **Guidance information of data source, and educational information of CT/EAP:** communicating and explaining myTomorrows services, as well as the intention for building such a search, to increase the transparency and accumulate trust. (#3.2, p.77)
7. **Source databases:** indicating the capability of myT search and communicating data sources as is what HCPs regarded. (#3.1, p.77)
8. **Map view:** enhancing the search by presenting an overview of search results visually by geolocations. (#2.12, p.77)
9. **Study design & related study results:** curating scattered information of published study results, to provide solid information that HCPs could make decisions upon. (#3.3, p.78)
10. **Treatment overview:** showing a merged overview of all developing treatments because the same intervention can be in use for multiple studies. (#3.3, p.78)
11. **Questionnaire-based patient search:** lowering the barrier for patients to run searches for their conditions and be aware of different options with detailed information. (#3.7, p.78)
12. **Relevance feedback collection:** motivating HCPs to provide explicit (positive and negative) relevance feedback on search results and improve the search system. (#4, p.79)

5.2.2 A visualized journey

With the visualized user journey in Search (Figure 49), the interactive search report remains an essential touchpoint for leading users to the search, as a part of myTomorrows services. After specifying the searching role, adaptive search interfaces will be presented to different searching roles. Patient search will be question-based, while HCPs will be presented with a regular (query-based) Search interface.

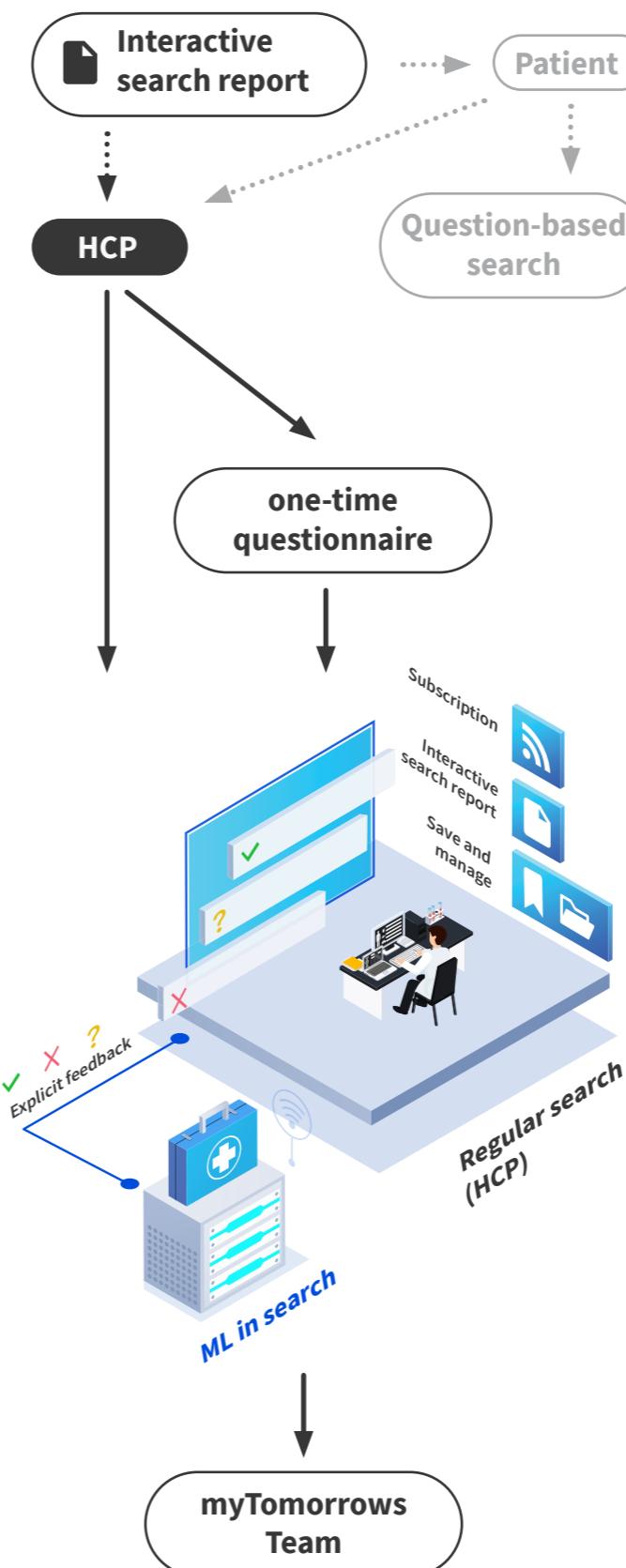


Figure 49. A visualized journe in search

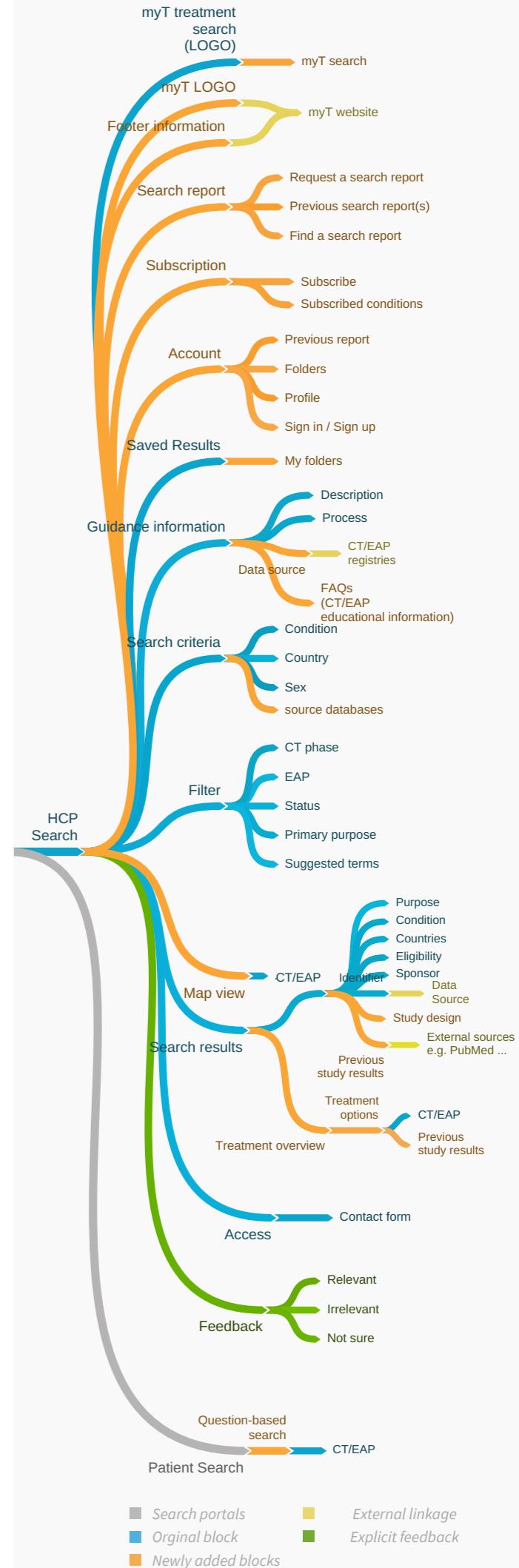


Figure 50. Reconstructed information architecture

Before the HCP Search launches, HCPs will be prompted with a one-time questionnaire for information on their professions. In the HCP search, HCPs can retrieve, review, and manipulate a search report(s) generated by myTomorrows. Besides, HCPs can choose to run searches themselves, subscribe to a condition, or save and manage search results. More importantly, feedback collection requests will be shown while HCPs interact with the system. The relevance feedback provided, together with the information of HCPs' profession, serves as the potential training data for system optimization.

At the end of the journey, HCPs will be guided to myTomorrows team for further steps if needed.

5.2.3 Five user flows

Based on the revised information architecture and experience journey, a low fidelity search interface with five primary user flows (p.89) was created in Figma (a UI/UX tool), serving as the basis for feedback ideas generation. The five primary user flows include one core flow of searching, reviewing results, and gaining access, and the rest are for four new features (#2, 3, 5 from the main changes in information architecture, p.87) in Search.

Core user flow

The core user flow starts with the initial entry point on the myT website, and then it leads users to run a search (search homepage), review search results (search engine result page & result content page), and contact myT team (contact page with a form) to access specific treatment options when needed. Along with the flow, users could take action on the search results (download, share, and print).

Subscription

Subscription allows users to stay updated (with alert) in the fields of interest, and there are two portals for subscribing to a condition (search query). One is to subscribe through the search engine result page, while the other is by adding a search query directly on the subscription page.

Once subscribed, users could set email alerts or create RSS feeds.

Search report

Search report aims to bring the current search report (PDF) online to enhance HCP engagement and increase myTomorrows search exposure. The flow starts with an email notifying users of a search report, and the email contains a code and a link to myT treatment search homepage. On the search homepage, users need to enter the search report page via the navigation bar of the interface, and then enter the code for retrieving the report. The search report will be added to the user account (if registered) once retrieved, and it is editable (adding or removing treatment options).

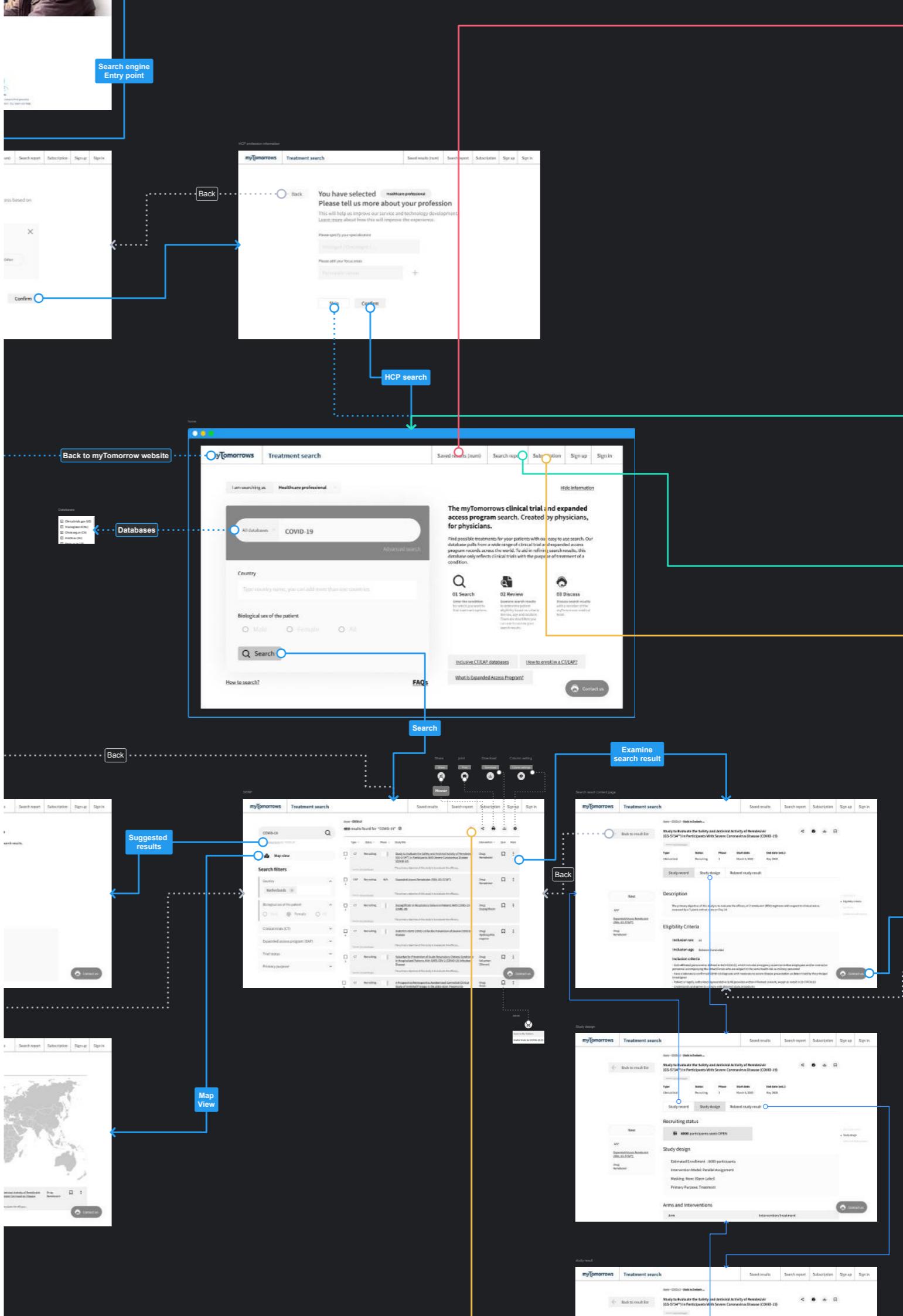
myFolders

It empowers users to manage saved search results, especially for internal HCPs, to reduce repetitive workload. In the saved results page, users can create folders for sorting saved results. Once a folder is created, users can save search results directly to a folder.

Suggest study results

Considering the technical boundaries of curating scattered study results from various sources, suggest study results aims at involving HCPs to link published results to a study result. Users can paste a link to a published paper and submit it. Before it is visible to other users, the submitted information needs an internal screening for information credibility and quality.

The complete wireframed search flow in Appendices (p.168-184)



5.3 Feedback ideas generation

Feedback ideas serve as a probe for investigating how different interactions and motivators affect user engagement in providing relevance feedback. Based on the design requirements (chapter 4, p.79), examples from the practices (p.69), and results from creative session (p.85), a total of nine feedback ideas (p. 94-96) were created with a broad spectrum and variety. Despite the difference in interactions, all ideas were built on the same feedback mechanism.

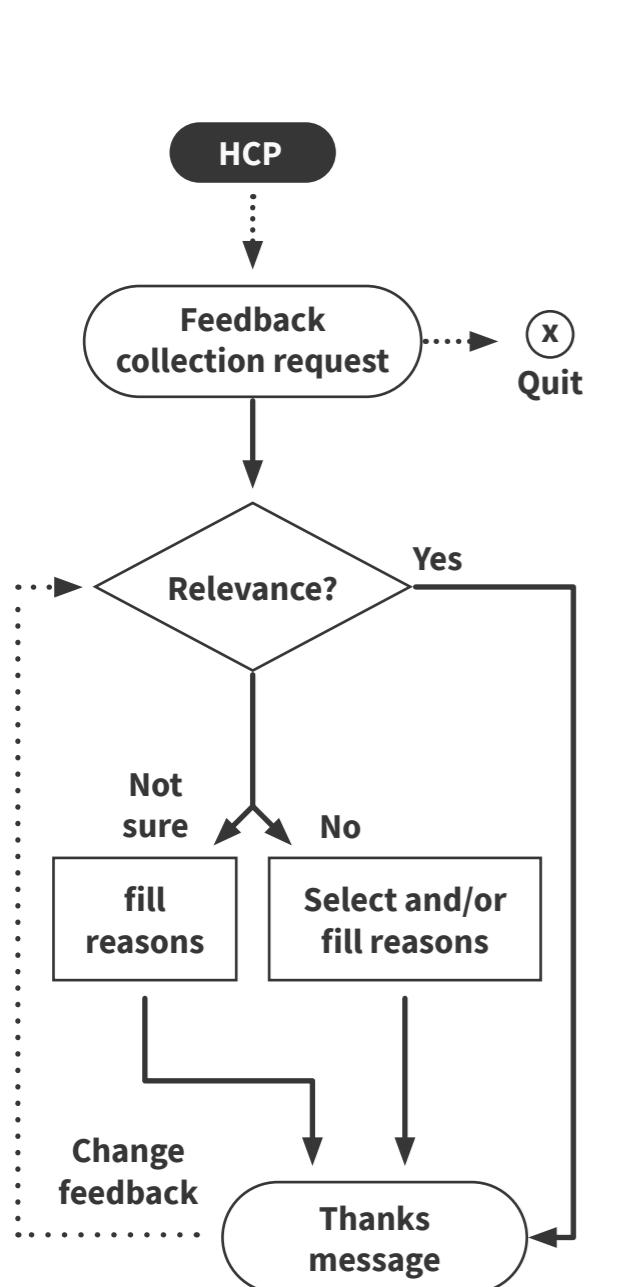


Figure 51. Flowchart of a standard feedback collection mechanism
(Drawn by author)

5.3.1 Feedback collection mechanism

The design of the feedback collection mechanism is a fusion of the patterns and findings discovered in previous research. Fundamentally, it uses binary judgment (feedback collection in practice, p.66), predefined options (relevance reasons, p.65), and free textbox to collect relevance judgment and underpinning reasons.

Figure 51 demonstrates the flow of relevance feedback provision. First, giving feedback is optional, and once users are in the flow, it starts with a question to solicit the relevance judgment with three predefined options: yes (positive), no(negative), and not sure (uncertain). A thanks message will follow after users make selections as a praise (p.41), but for the no and not sure options, there is one more step where a list of reasons and a free textbox will be presented for users to select and (or) fill. Besides, users can modify the feedback when they change their minds or want to fix a mistake because relevance changes over time (Mizzaro, 1998, p. 36).

Figure 52 shows an example of a relevance feedback request UI according to the feedback collection mechanism.

5.3.2 Nine feedback ideas

Since relevance judgment could be made at various stages in a search flow (p.63), thus, all ideas generated were primarily based on the information-seeking process by stage (p.30) and inspired by examples previously gathered.

Relevance?

Is this what you are looking for?

Yes

No

Not sure



Select and (or) fill reasons

Tell us why you choose:

No



It is irrelevant to the search keyword

The filters didn't work

It will not benefit my patient

The study is poorly designed

Other

Confirm

Thanks message

Thank you for your contribution! This helps improve our system for better search experience

[Change the feedback](#)



Figure 52. An example of feedback request

By category, all ideas could be classed as pre-screening relevance feedback, In situ relevance feedback, and post hoc relevance feedback (Table 12). Pre-screening relevance feedback refers to those collected while users are pre-screening the results in the search engine result page. In-situ relevance feedback means feedback collected while users are reviewing the content of a search result(s). Lastly, post hoc relevance feedback represents feedback collected after users have finished reviewing a search result(s) (e.g. while contacting for access).

In addition to the stages, different motivators also blend in those ideas. Idea 7 uses challenge, idea 8 tries to turn feedback provision as a utility tool, and the last idea (9) employs gamification elements for personal enjoyment. Page 94-96 lists all feedback ideas with UI.

Stage category	Idea	Motivator
Pre-screening feedback	Idea 8	Utility
	Idea 1	Altrism
	Idea 2	Altrism
	Idea 3	Altrism
	Idea 4	Altrism
	Idea 5	Altrism
In situ feedback	Idea 9	Enjoyment
	Idea 6	Altrism
Post hoc feedback	Idea 7	Challenge

Table 12. List of design requirements aimed at increasing perceived trust

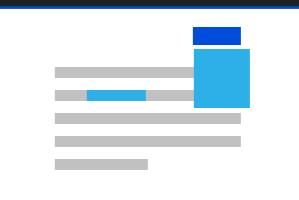
This screenshot shows the myTomorrows Treatment search interface. At the bottom of the page, there is a feedback collection pop-up with three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



1) Bottom pop-up

The feedback collection request slides up from the bottom of the screen while users are reviewing a search result. It is common in web experiences, and it attracts users' attention to a large extent. However, it might tend to interrupt users.

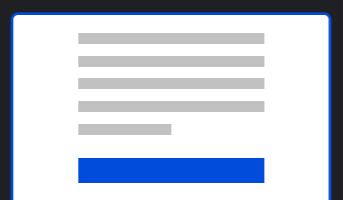
This screenshot shows the myTomorrows Treatment search interface. A feedback collection pop-up is displayed at the bottom of the page, containing three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



4) Action + 'Pointing out'

This idea is similar to the third one, but it differs in the labels (Useful, Useless, Not sure), icon (emotional faces) and the functionality, which allows users to select content elements (applicable only to negative or uncertain feedback) so that the system could identify on what the judgment is based (p.69).

This screenshot shows the myTomorrows Treatment search interface. At the bottom of the page, there is a feedback collection pop-up with three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



2) Bottom content

Feedback collection request stays at the bottom, as a part of the content. It is another typical design in web experiences, and it is always visible to users. More importantly, it is less obtrusive than the first idea.

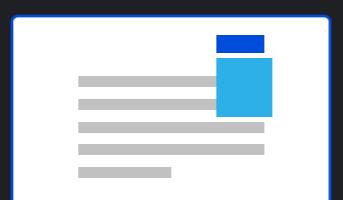
This screenshot shows the myTomorrows Treatment search interface. A feedback collection pop-up is displayed at the bottom of the page, containing three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



5) Pop-up comparison

When a user has reviewed more than one search result, there will be a pop-up from the bottom, asking users to compare which one is more relevant. This idea tries to experiment with a novel approach to collecting relevance feedback.

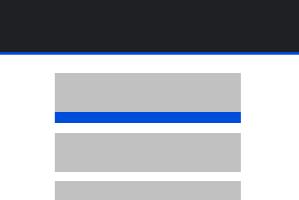
This screenshot shows the myTomorrows Treatment search interface. At the bottom of the page, there is a feedback collection pop-up with three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



3) Action + Selection

Feedback collection request appears as a button with a 'feedback' label, and it sits next to other action buttons. Each option accompanies an icon (thumb up/down), and they will show up when the cursor is hovering on the button. This idea aims to minimize the interruption that users may experience in the previous two.

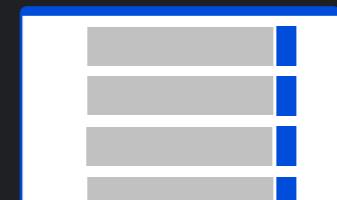
This screenshot shows the myTomorrows Treatment search interface. A feedback collection pop-up is displayed at the bottom of the page, containing three buttons: 'Yes', 'No', and 'Not sure'. The main content above the pop-up is a study record for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'.



6) Surrogate message

When users finish reviewing a search result and return to the search result page, the request will appear below the reviewed search result. This idea attempts to remind users to provide feedback on each reviewed result with the confidence that it does not tend to interrupt users reading.

The screenshot shows a search results page for 'Treatment search' with the query 'COVID-19'. It lists several clinical trials, each with details like Type (e.g., CT, EAP), Status, Phase, Title, and a brief description. To the right of the list are three blue rectangular feedback bins.



7) Task session

When users have reviewed a few search results (5 for instance), a message asking for feedback appears on the top, directing to a page with a list of reviewed results. This idea tries to present feedback provision as a task or challenge, trying to trigger users' intrinsic motivation.

The screenshot shows a search results page for 'COVID-19' with a large number of results (400). The interface includes search filters on the left and three blue feedback bins on the right side of the results list.



8) Feedback bins

There are three feedback bins (promising bin, trash bin, question bin) fixed to the right edge of the screen, allowing users to sort search results by dragging/dropping. Each bin will display an accumulated number of sorted results. This idea tries to turn feedback provision into a tool that helps users manage search results.

The screenshot shows a detailed view of a clinical trial titled 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir [GS-5734™] in Participants With Severe Coronavirus Disease (COVID-19)'. It includes sections for 'Description', 'Eligibility Criteria', and a 'Feedback bot' section at the bottom right.



9) Feedback bot

A gamified feedback bot asks users the relevance while reviewing search results, and it grows as users provide feedback. This idea incorporates common gamification elements in the health domain: avatar and levels. (Johnson et al., 2016), which may induce personal enjoyment.

5.4 Usability testing and ideas speed dating

Sessions of usability testing of the baseline search and feedback ideas speed dating (Davidoff et al., 2007) were conducted for three objectives. First, benchmark the UX of the baseline search and discover potential usability issues. Second, prioritize the four new features. Lastly, gain insights into explicit relevance feedback collection for forming the basis of feedback conceptualization.

5.4.1 Methods

A total of seven participants joined the sessions, including three peers (Dfl) and four employees of myTomorrows, three of whom were internal HCPs and one from the business department. The selection of test participants was based on the two criteria, professional design background with no medical background and internal employees with medical-related education. In this case, those with design background could provide more professional feedback on the overall user experience. The other group could give more in-depth suggestions due to their familiarity with the medical context and myTomorrows Search.

Those sessions (Figure 54) were carried out individually with participants (via ZOOM or Microsoft Teams), following the same protocol comprising three testing sections (as illustrated below): **core flow comparison, new features, and feedback ideas speed date**. Before the start of each section, the participant was presented with a fictional scenario for better immersing the context. They were asked to think aloud throughout the session, and a questionnaire was followed after the completion of each section to investigate their experiences. All sessions were video-recorded as a part of the data collection.

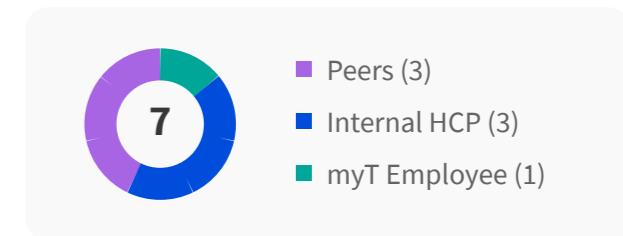


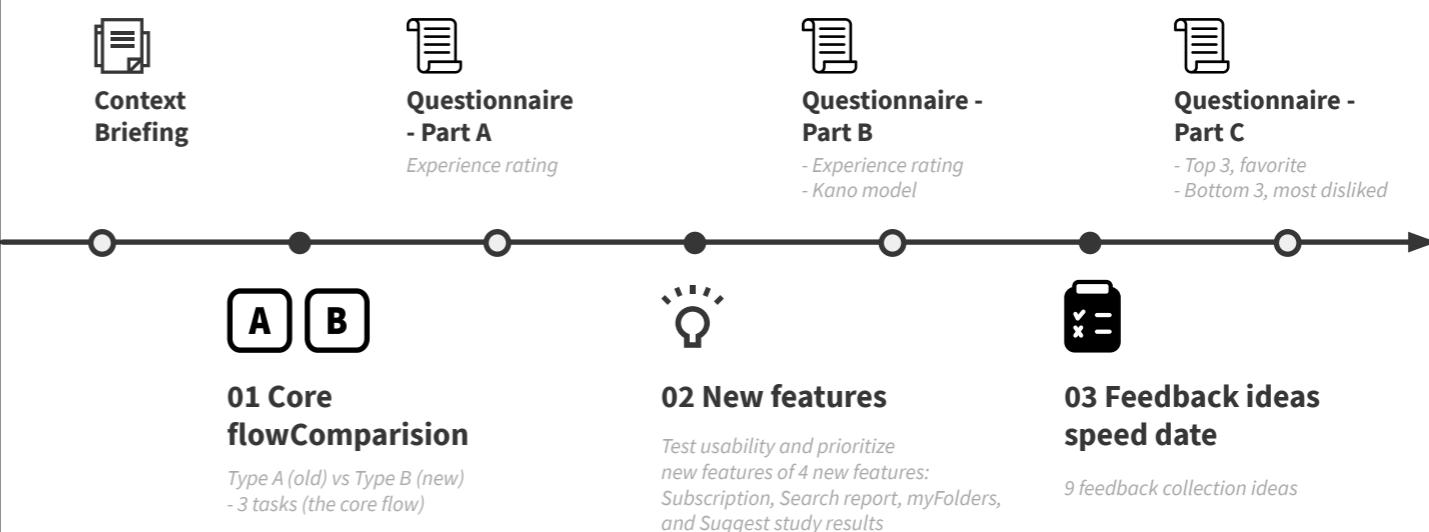
Figure 53. participants distribution

Fictional scenario:

Your role: You are a doctor working at Erasmus MC

Your patient: 52 years old Dutch women, diagnosed with COVID-19

You have a patient tested positive with COVID-19 in severe condition. As there is no specific treatment, you decided to try out the 'Treatment search engine' you heard of before. You knew that the company behind provides services to assist you in enrolling your patient(s) to a particular clinical trial or expanded access program. Now you are on this search engine, and you want to find a treatment option(s) for the patient.



01 Core flowComparison

Type A (old) vs Type B (new)
- 3 tasks (the core flow)

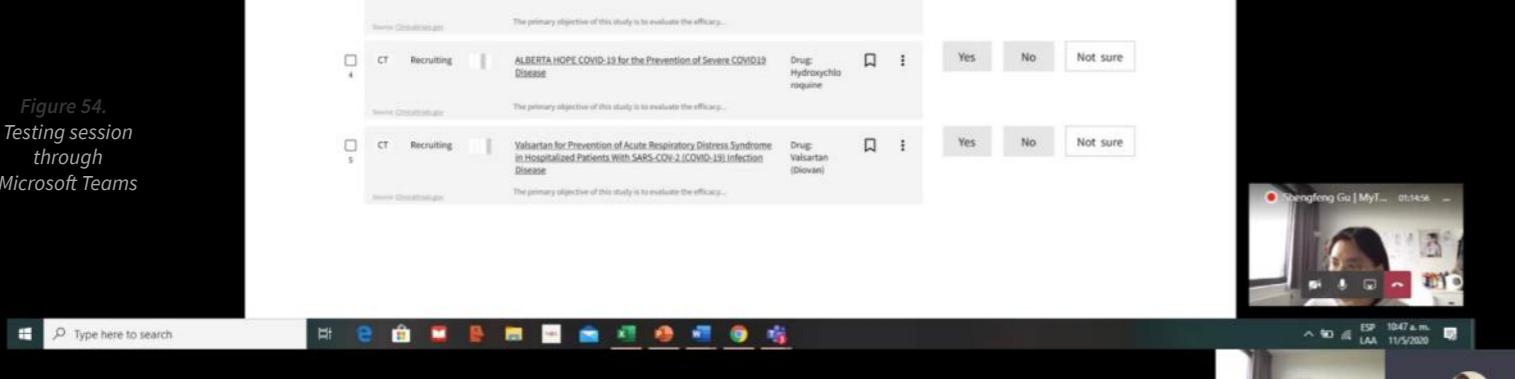
02 New features

Test usability and prioritize new features of 4 new features:
Subscription, Search report, myFolders, and Suggest study results

03 Feedback ideas speed date

9 feedback collection ideas

Figure 54.
Testing session
through
Microsoft Teams



01 Comparison

The first test section aimed to benchmark the new design concept (Type B) by comparing the core user flow to the original external Search. A wireframe of the original design (Type A) was made in Figma to mitigate the gap between a design prototype and a real product. Besides, the testing sequence of Type A and B were randomized to avoid anchoring bias (Furnham & Boo, 2011). Participants were asked to perform three identical tasks with the two versions:

1. Run a search based on the fictional scenario and examine the first two results
2. Save the first two search results and review all saved results
3. Contact myTomorrows for accessing a specific treatment (e.g. Remdesivir EAP)

02 New features

This focused on testing the user flows and desirability of the four new features. There were three tasks in this section, but myTomorrows participants were assigned to perform an extra task because it requires professional medical knowledge:

1. Subscribe to the condition and set an email alert
2. Review and edit a search report
3. Create a folder for saved search results
4. Add a newly published study result to a search result (internal participants only).

03 Feedback ideas speed date

The last section presented participants with nine relevance feedback collection ideas (p.92-94), serving as a design probe to elicit users' opinions and preferences on providing explicit feedback. Participants have the freedom to provide feedback or not as their wishes.

Questionnaire

The questionnaire comprised three parts following the three test sections. Part A focused on the rating of the experience of Type A and Type B for comparison, and Part B collects the ratings of each new feature and its usefulness with a Likert scale. The last part asked participants to vote feedback ideas with eight votes for the top three, bottom three, favorite, and the most disliked one.

5.4.2 Study results

Baseline search experience

The results from the questionnaire (appendices, p.185-194), notes of heuristics, and transcribed screen records were used as the source of insights. For the first test section (comparison), the overall experience of the baseline search design slightly outperforms the original one. All internal HCPs preferred the new design because of the completeness, richness of information and additional features that streamline their workflows, such as myFolders.

However, the new design was perceived as visually crowded by the rest of the participants, especially on the search homepage. It might be the consequence of displaying more information and roughness in visual design. More importantly, the design of gaining access to a study in the baseline search has a significant flaw that users would regard the 'Contact us' button as general customer support, instead of the entry point for access.

Affinities classified all study results into four categories (Table 13) of impression, UX complaints, suggestions, and facts.

Impression	<p>(+ positive)</p> <ul style="list-style-type: none"> More clear to save and find saved results Feels clean and information is organized (internal HCPs) More informative The map view helps identify studies easier Has a better review experience (compared to the original one) <p>(- negative)</p> <ul style="list-style-type: none"> The menu on the top is not eye-catching The interface is visually crowded
UX complaints	<ul style="list-style-type: none"> Positions of myTomorrows LOGO and treatment search are confusing Collapsed filters are indirect Usability issues of the checkbox on the left to search results Gaining access to a study is indirect and unclear Study record covers the meaning of study design Lacks of confirmation after filling out the contact form 'Related study result' does not fully match with its content
Suggestions	<ul style="list-style-type: none"> Suggest bigger font of the study title (content page) 'Next' might be unnecessary (content page) Suggests adding the corresponding icon to saved results (navigation bar) Wants to have data (study results) visualization and overview Contact for access appear in per trial Needs more clear communication or guidance of next steps (for gaining access) Expects detailed location information including contacts Wants to easy access to interesting study results
Facts	<ul style="list-style-type: none"> CT.gov has higher search recall myT participants prefer CT.gov for access because of directness Assumes that the higher the result is, positioned the most relevant it is

Table 13. List of findings of the search concept by categories

New features

Subscription

The average score of this feature was 3.5 (out of 5, Std=1.28), and most participants (6/7) were able to achieve the tasks assigned to them. Participants had very distinctive expectations due to their previous experience and the ambiguity in communication (i.e. wording). Some thought it works similar to newsletters; some thought it was technically not feasible, and one participant was used to save the web page as a way to keep following updates manually. Besides, there are some UX complaints with the feature as follows:

- Not clear and requires more explanation
- No confirmation when subscribed
- Email alert (text) is too small

Search report

The experience rating was rather low, with an average of 2.85 (out of 5, Std=1.18). The reasons were mainly the usability issues participants encountered, resulting in a low success rate (1/7) of adding a search result to the report. Moreover, external participants felt that the purpose of the search report was not clearly communicated. Besides, two internal participants mentioned that it did not feel like a search report because of the lack of summarization and other supporting

information (e.g. the creator of the report). Additionally, there are some UX complaints with the feature as follows:

- Edit and save changes is unnecessary
- Checkbox is confusing
- The link from email does not direct to the final page
- The search bar on the home page can not search the report
- The thumbnail of the search report while adding is not obvious

myFolders

The experience score of myFolders was moderate (Mean=3.14/5, Std=0.99). All participants succeeded in the task but encountered some minor issues while removing results from a folder. In general, it worked but did not closely match the users' mental model:

- Moving saved results to a folder is confusing
- 'Learn more' while creating a folder is unexpected
- Removed result from a folder should be back restored in the saved results page
- Expects to have a place to view removed results in the folder
- odd to see a blank page in saved results when some are saved in folders.

Suggest related results

This task was easy to achieve as all (myT) participants succeeded in the task, and there were not any usability complaints in it. In general, internal participants had a positive attitude (with an average score of 3.5/5, Std=1.12) towards engaging users to build the search by suggesting published study results, because such information is an essential criterion for HCPs to make decisions. However, they were concerned that it would bring more burdens to external HCPs for doing this. Also, they would assume that study result data is supposed to be provided and ready by the search system or its affiliated organization.

New feature prioritization

KANO model (Kano, 1984) was employed to prioritize the four new features in the baseline

search design. The model classifies features into four categories (Figure 55) and prioritizes them with the sequence of must-be, performance, attractive, and indifference.

The results (Figure 56) suggested that 'search report' is considered on the top as a performance feature, followed by 'myFolders' and 'suggested related results' (Attractive feature). However, 'subscription' (indifference feature) was the least appealing feature to participants, mainly because of its heterogeneous expectations participants had.

Overall, the prioritization is in accordance with the current service funnel (p.20) where the search report plays a vital role in business conversion, and the rest are useful add-ons to the search.

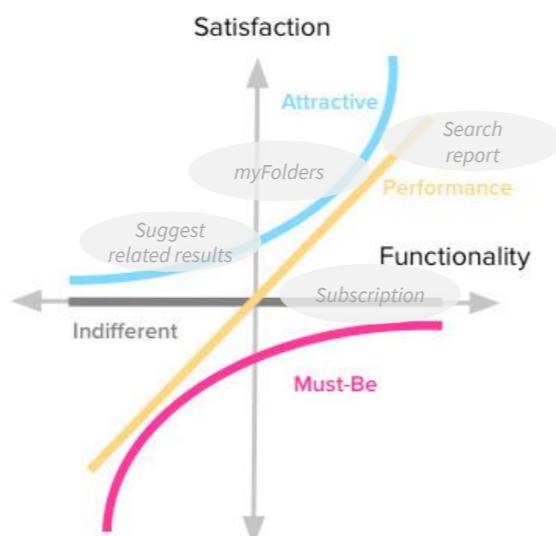


Figure 55. Visualized feature categorization in Kano model (KANO, year)



Figure 56. Feature prioritization (drawn by author)

Speed date results

Participants had three votes for the top three, another three for bottom three, one for the favorite, and one for the most disliked feedback idea (Figure 57). For the favorite votes, the most prominent one is the first idea (Bottom pop-up, p.92) with three favorites plus one most disliked, followed by the sixth idea (Surrogate message, p.93) with two favorites. On the contrary, the seventh idea (task session, p.94) is the most disliked one with three votes. However, the last two ideas remained controversial as they received a similar amount of positive and negative votes.

The reasons for the results varied from person to person. Nevertheless, to the most disliked idea (seventh), participants had the consensus that the message is not inviting, and it would take too much effort to provide feedback.

Feedback 1 - Bottom pop-up

This idea had the highest favorite votes of 3 as well as 1 most disliked vote. Participants liked it because it is direct, easy, and they are familiar with this kind of interaction in Web experience. Most participants thought it doesn't interrupt reading, however, it could be perceived as annoying because of the unpredictable moment of popping up.

P2, "The 1st is my fav, because it doesn't affect my reading experience that much, and I can give feedback whenever I want, I don't need to search for it."

P5, "So that...Well, I dislike the 1st one, the worst because then it's near the screen all the time, then you have to click it away. That's really annoying."

Feedback 1 - Bottom content

Most (4/7) participants liked it because of the easiness of giving feedback and familiarity with such a position. More importantly, it did not annoy users less compared to the first concept. However, it was mentioned that it could be easily missed because users would not even scroll down to the bottom in some cases.

P6, "I didn't like the bottom, but I didn't...is that this is not that I don't like it. But I think it's very easy to miss this if you don't scroll down until the end of the page."

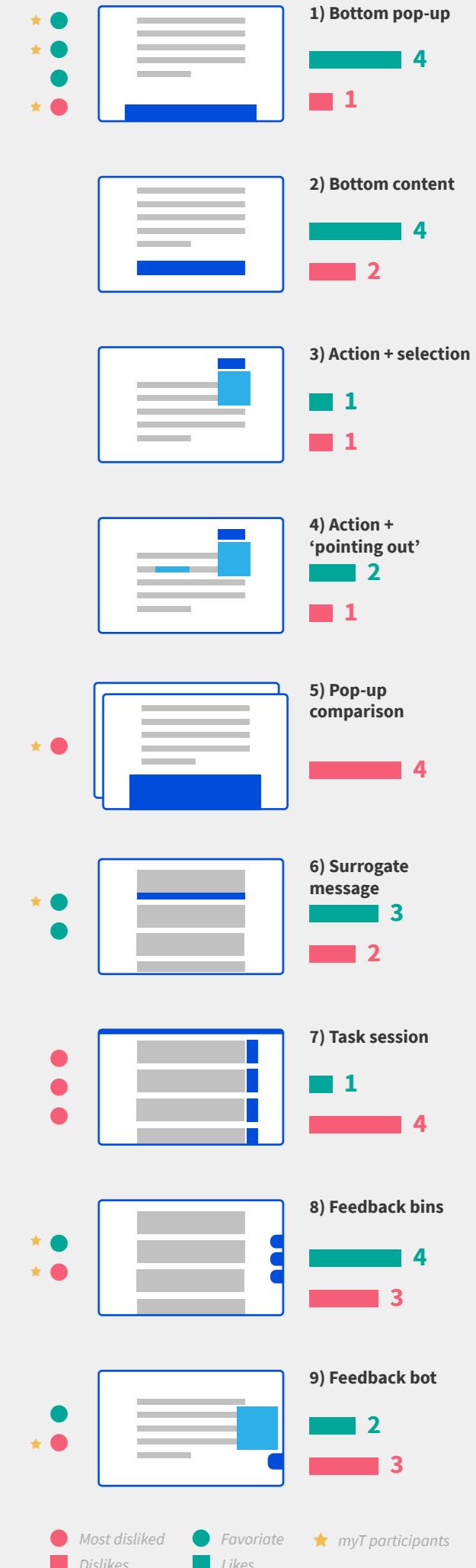


Figure 57. results of voting (drawn by author)

Feedback 3 - Action + selection

Participants did not expect feedback to be 'hidden', and they did not want to look for the place to give feedback. Even though it was easy to learn where to find it, participants perceived an extra step to give feedback as they reckoned a click on the button is needed before making selections. The reality was that the action button would automatically display all options once the cursor is hovering on the feedback button. There is no extra step for users to interact, but the cognitive step in the head.

P4, "I wouldn't, I wouldn't see this. But I like the icons."

Feedback 4 - Action + 'pointing out'

Participants had a similar experience as the previous one, but the wording 'useless' was perceived more aggressive. The free text box and free selection of web elements were liked, as well as icons that helped users quickly identify options. It indicates that participants prefer having more control and a bit more novelty in interactions, but it does not change their overall perception of providing feedback as an additional task.

P6, "Okay, so I like better relevant or irrelevant. I don't know. 'Useless' sounds very aggressive. But I like that you can give your feedback directly here."

P7, "I think it's nice that people can write, you know, whenever they think about it."

Feedback 5 - Pop-up comparison

All participants had a negative experience with this idea because the question was regarded as indirect, unclear, and effort-consuming. They assumed that it was supposed to be the system to make comparisons. Also, it was mentioned that it occupied too much screen space visually.

P3, "It feels a bit strange, and it's hard to compare...will take quite a lot of efforts."

P4, "It's quite a strange question."

Feedback 6 - Surrogate message

Around half (3) of the participants thought this idea just right as it was not annoying, easy to

respond, and appeared in a reasonable place. However, some had concerns about its potential visual burdens when users are reluctant always to provide feedback, and the requests will accumulate. Also, one participant expressed his dissatisfaction that it interrupts the flow before reviewing the next result, and it should not drag users back to what should have been done in the previous review.

P1, "I think it's pretty good because it's quite reasonable to be here after reviewing. It triggers me to click it away."

P4, "Because if you don't want to open them, they will be here unless you click it away. Yeah. And that would be an extra task."

Feedback 7 - Task session

Overall, This idea was the most disliked one because of the perceived cognitive burden and indirectness. Participants explicitly said that they would not even click the message because the message did not catch their attention and has nothing to do with their goal of finding relevant information. Besides, one participant said he would not even read such a message as it resembles the cookies for user consent.

P1, "Oh, I didn't see this, give feedback now. It's not that obvious."

P4, "I wouldn't do this because it's so much effort... Because you have to review them several times. Yeah, there are several and one so it takes a lot more time than at one page."

Feedback 8 - Feedback bins

There was no consensus on this idea among participants. For those who liked, it offered an additional function while giving feedback, and could streamline their workflow. In comparison, the rest was frustrated by the usability (drag and drop), the usefulness of the function.

P4, "because it (concept 8 - bins) would clash with the save options. And it's clear it's a feedback loop that is in there."

P6, "yeah, I really like that it is very simple to use if you see the icons are these things immediately and you can select the trials and put them on the bins, I think it is very simple very fast."

Feedback 3 - Feedback bot

Similar to the previous one, this idea was controversial as well. Nearly all (3) internal participants disliked this concept because it was regarded as not fitting in such a serious context. However, one participant (peer) felt it has a human touch in a serious context, and it could ease the tension in a serious context.

P3, "It has a human touch with end-users. The search is very serious and technical... But if I were a doctor, maybe I would like the 3rd concept."

P4, "Because the subject is really sensitive and it's research. I wouldn't add such a thing that's a little bit more playful, so it's kind of strange to have it in this context."

2) Visibility (of the feedback request)

Participants would not want to respond to feedback collection requests actively, let alone looking for them. Therefore, how visible such requests are to users is vital, as revealed through responses in ideas two and three. However, too visible might be bothersome to users, for instance, the fifth idea that occupies a significant portion on the screen. Thus, keeping the balance of being easily visible and not over eye-catching is essential. It has much to do with timing and its embodiment.

3) Timing (of showing up)

When feedback requests would appear matters much because the wrong moment to show the feedback request to users would be a distraction from users' search flow. Besides, participants expressed that authentic relevance feedback could only be given after finishing the content examination. Therefore, being able to distinguish the moment when users have finished reading is vital. It was also why the first, second, and sixth ideas were liked because it asks the participants to provide feedback at the end of a review. However, the first idea failed to accurately distinguish the moment that users have finished reading, which was why it received one vote of most disliked.

4) Embodiment (of the feedback request)

As a part of the graphical user interface, the embodiment of feedback collection request has a significant impact on participants: the size, the position, iconography, and similarity to other experiences. Participants mentioned that due to the screen size idea five occupies, they felt annoyed and distracted. Also, the second concept was somehow 'invisible' because users do not read through the whole page every time, especially when the content could be judged as irrelevant from the beginning of the page. Besides, iconography could help participants quickly recognize the meaning and intention of feedback requests (i.e. idea two and three). Furthermore, the embodiment (idea 7) could be associated with other experiences that users regularly ignore, such as GDPR consent pop-ups (Schofield, 2019).

5) Perceived effort (of the feedback request)

Perceived effort links to the embodiment. It could be sensed from the number of feedback requests users see at once, as well as the complexity of and familiarity with questions and interactions. Users preferred feedback provision to be direct and straightforward. For instance, the fifth idea asked users to compare two results, requiring more cognitive effort to recall the previous and compare. Nevertheless, providing feedback on the seventh idea is as easy as many other ideas, but it was presented with a list of requests and required an extra step to enter. Thus, participants felt an intensive workload. The same as the worries participants had on the sixth idea that feedback requests accumulate if users do not respond, and it would be an extra visual burden.

6) Control (over the feedback)

The controls participants have on relevance feedback requests influence the experience. Participants preferred it to have more controls, such as a free textbox or free selections of elements (in idea four). However, having more control does not necessarily mean that they would be more engaged in relevance feedback provision, but it is

worth investing. Moreover, when participants do not have specific controls over a feedback request (e.g. dismiss or skip a feedback request), they might be irritated to some extent.

7) Tone (of the feedback request)

Participants pointed out that using different tones of asking for relevance feedback would bring perceptual differences. In ideas three and four, the former uses 'relevant' and 'irrelevant' as the options for (positive and negative) relevance, while the latter uses 'useful' and 'useless'. One participant had a worse experience with the latter because 'useless' sounded more aggressive than 'irrelevant' that sounds neutral.

8) Context fitness(of the feedback request)

The first eight feedback ideas are all technical without any embedding human touch. In comparison, the last idea tries to approach users with a humanized feedback request where a bot asks for relevance feedback. Nevertheless, most participants expressed the concerns of its fitness in such a serious context for healthcare professionals, and they felt it might be childish.

5.5 Feedback enhanced search conceptualization

Another brainstorming session with peers was hosted to generate more ideas on feedback collection interactions. The results, together with the findings and insights from the previous section, led to a feedback enhanced Search with three relevance feedback collection concepts using motivators of utility, altruism, and enjoyment.

5.5.1 Brainstorming session on feedback collection

A brainstorming session was hosted to generate more interaction ideas for collecting explicit relevance feedback, mainly focusing on designing motivations.

Methods

Three peer design students, respectively, from three Master tracks (Dfl, SPD, IPD) of IDE (TU Delft) were recruited for the session. The recruitment of the participants required their professional knowledge in design, and unfamiliarity with the topic.

The brainstorming session took place in Zoom (Video conference software, Figure 59) and Miro (Online collaboration tool). The session included three sections (Figure 58) and lasted for 50 minutes. It started with a twenty minutes briefing of the project context and objectives. Then, the core search flow (p.91) was demonstrated, and nine feedback interaction ideas (p.94-96) were shown to participants to help them form a better understanding of explicit relevance feedback

collection and avoid generating explored ideas.

The remaining thirty minutes were divided into two rounds of ideation (How-tos). Participants were asked to brainstorm feedback collection ideas for the search interface previously shown to them, with a special focus on motivations. The first round lasted fifteen minutes in total, where ten minutes were for drawing quick ideas, followed by a five-minute introduction to their ideas. The other round followed the same procedure as the previous one. However, the drawing part was shortened to five minutes to encourage participants to reflect more on their ideas.

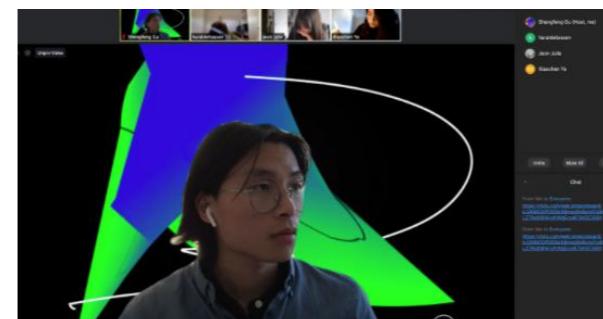
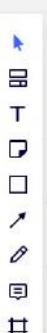


Figure 59. Screenshot of participants drawing (on Zoom)



The screenshot shows a Miro board with several sections:

- Context:** Includes screenshots of user interfaces and a section titled "ideas that have been tried and tested".
- Your show time!**: A section for presenting ideas, with three cards labeled P1, P2, and P3.
- examining**: A section with cards for "Marking", "Sorting", and "Rating".
- Post session actions**: A section with cards for "Marker (notes)", "Rating", and "Reminder".
- Motivations**: A section with cards for "giving", "unlocking", "material reward", and "utility".

Figure 60. Screenshot of the brainstorming session (on Miro)

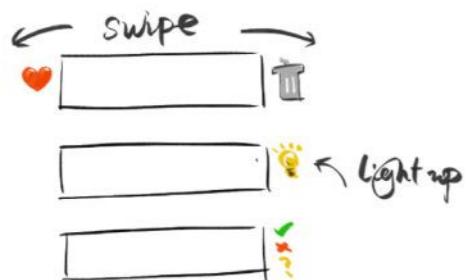
5.5.2 Feedback ideas by stage

A total of nineteen ideas (including twelve for feedback collection and seven focused on motivating users) were gathered from participants. All ideas fit in the categorization by stages in search (Table 12, p.93). Surprisingly, most feedback collection ideas were around the stages of reviewing the search result list (7), which is in accordance with the pre-screening feedback. In comparison, the rest is on examining the content (3, In situ feedback) and taking actions on search results (4, Post hoc feedback).

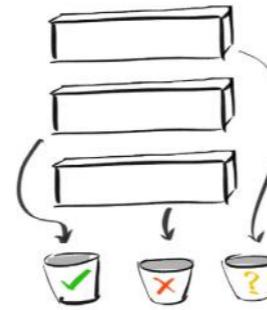
Pre-screening feedback

Most ideas appeared at this stage. By different types of interactions, those seven ideas formed four clusters of marking (3), sorting (2), ranking (1), and filtering (1).

Marking (3 ideas): there were three ideas in this cluster, and they borrowed ideas from real life (i.e. using stickers and lighting bulbs to highlight search results), or other popular digital experiences such as Tinder (A popular dating APP) swipes.

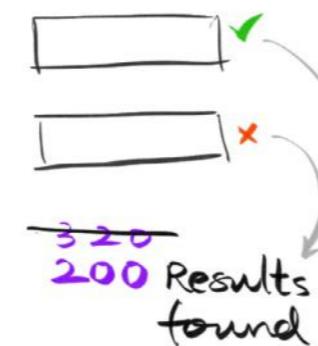


Sorting (2 ideas): sorting shares the same fundamental idea with marking, but the main difference is its capability of gathering sorted results. Both two ideas used a similar analogy of desktop experience of folders.



Ranking (1 idea): ranking enables users to rerank search results in an order that meets a user's information need.

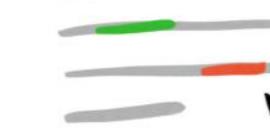
Filtering (1 idea): filtering search results by using feedback to optimize search results immediately.



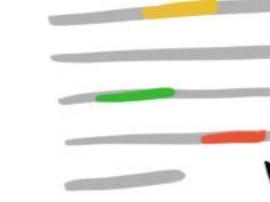
In situ feedback

Merely three ideas focused on collecting relevance feedback while users are reviewing the content of a search result. However, each idea is distinctive in terms of interactions.

Rating: this idea is placed as part of the content, asking users to provide feedback with a scale for ratings on different aspects of relevance.



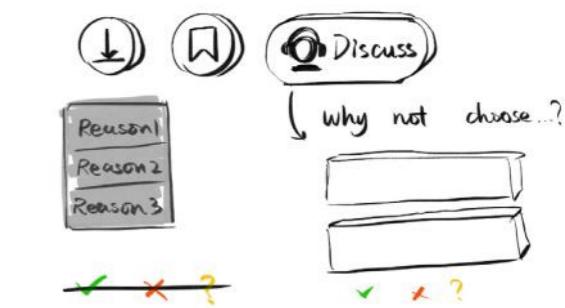
Marker: the marker idea allows users to highlight the interesting parts of the content as an indication for content specific relevance.



Reminder: this idea asks users to leave relevance feedback before returning to the search result list page.



specific study, it implies that they are interested in these results, and the system has helped them to some extent. Therefore, two ideas were about asking users for the favor of providing relevance feedback when gaining access to a study. The generation of such ideas proved the principle of reciprocity (Table 2, p.43).



Motivations for feedback provision

Five ideas were gathered on motivating users to provide feedback, and they employ motivations of personal enjoyment, reward, and utility.

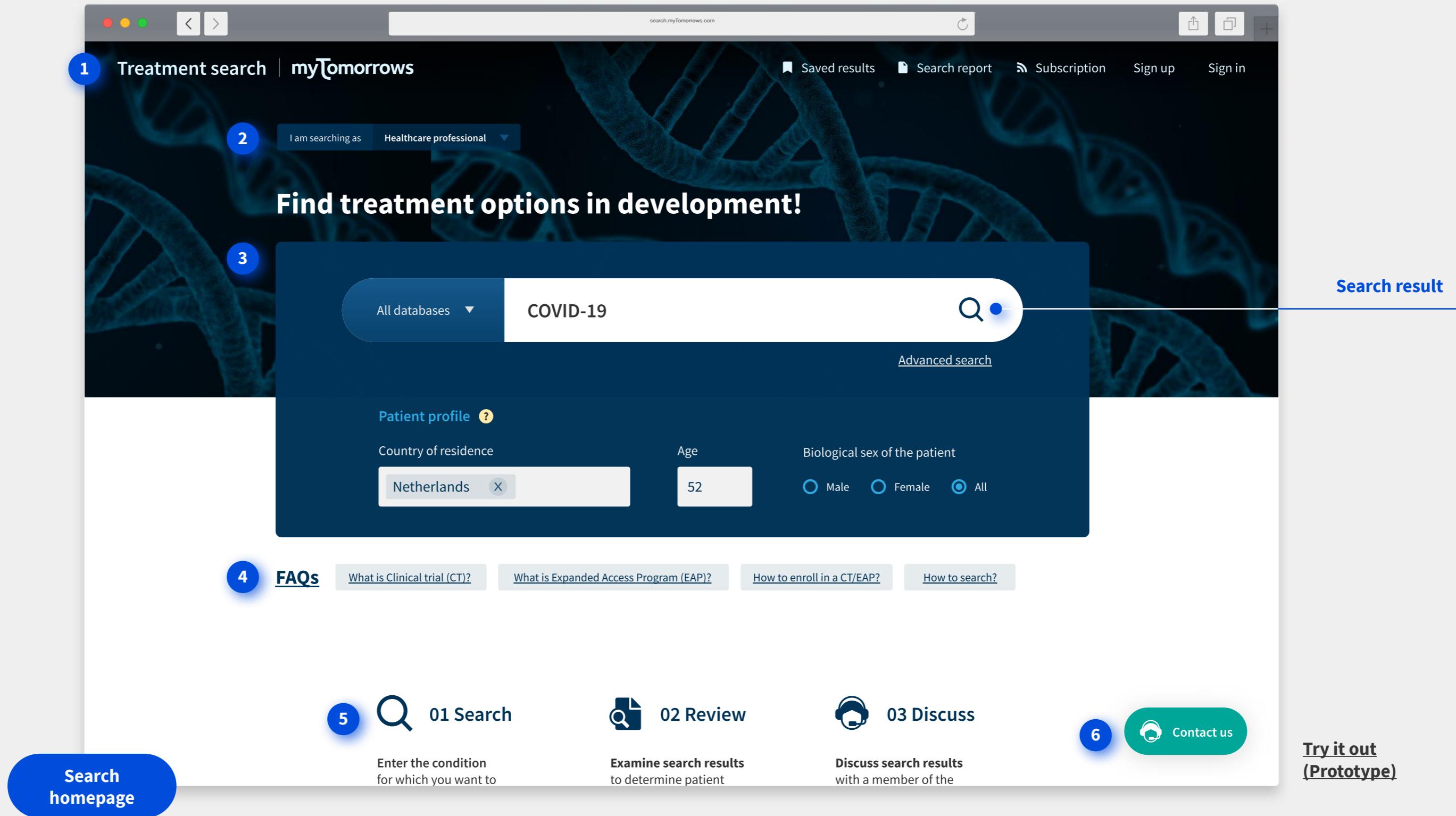
Personal enjoyment (3 ideas): three ideas focused on increasing personal enjoyment of providing feedback. Two were about showing users a progress bar to visualize the accumulation of feedback provided, while the other was about a highly responsive feedback mechanism allowing users to click likes for multiple times.

Reward (2 ideas): there were two types of rewards participants drawn, material reward (voucher) and virtual reward (unlocking more features in search as feedback provision accumulates).

Utility (3 ideas): two ideas were about displaying users' records of search behavior and dwell time in the search.

Post hoc feedback

While users are taking actions on search results such as download, save, or request access to a



5.5.3 Iterated baseline search

The findings and insights from the previous user testing with peers and internal employees supported the iteration of the baseline search (p.89) to resolve the identified complaints on

usability, incorporate suggestions. Besides, a visual design adapted from the myTomorrows visual imagery was applied to increase the fidelity of the baseline search design concept. Table 19 (Appendices, p.205) shows a list of changes from the low fidelity version (p.89).

1. Navigation and menu
2. Searching role switch
3. Search area
4. FAQs and educational information on CT/EAP
5. Instructions and additional information
6. Contact center

The image above shows the homepage of the baseline search concept. The iterated design concept focuses much on emphasizing the search area (3) as well as instructional information (4 & 5) around myTomorrows services and educational information on Clinical Trials (CTs) and Expanded Access Programs (EAPs).

The screenshot shows the myTomorrows Treatment search interface. At the top left is the search bar with the query "COVID-19". To the right are navigation links: "Saved results", "Search report", "Subscription", "Sign up", and "Sign in". On the far right, there are "Subscribe" and "View Setting" buttons, and a "Unranked" dropdown menu.

Left Column (Search Criteria and Filters):

- Search Bar:** COVID-19 (3)
- Map View:** Map view (4)
- Patient Profile:**
 - Country of residence: Netherlands (5)
 - Age: 52
 - Biological sex of the patient: Male, Female, All
- Search Filters:**
 - Result types: Clinical Trials (CT), Expanded Access Program (EAP) (both checked)

Right Column (Search Results):

400 results found for "COVID-19". The results are categorized by intervention type: "All studies" (selected) and "Intervention overview".

Rank	Intervention Type	Status	Phase	Study Title	Drug	Action Buttons
1	CT	Recruiting	III	Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)	Remdesivir	
2	EAP	Recruiting	N/A	Expanded Access Remdesivir (RDV; GS-5734™)	Remdesivir	
3	CT	Recruiting	III	Dapagliflozin in Respiratory Failure in Patients With COVID-19 (DARE-19)	Hydroxychloroquine	
4	CT	Recruiting	III	ALBERTA HOPE COVID-19 for the Prevention of Severe COVID19 Disease	Hydroxychloroquine	
	CT	Recruiting	III	Valsartan for Prevention of Acute Respiratory Distress	Drug:	

Bottom Right: Contact us button.

Search result page

1. Search result list with actions
2. Search result selection bar
3. Search bar
4. Map view of all studies
5. Search filters

The image above shows a standardized search result page (SERP). This page keeps the navigation bar on the top the same as the homepage, and it comprises two primary columns of information. On the left are the search criteria, filters, and map view option. The right column contains all search results and supporting information (e.g. search

result size and ordering mechanism). In addition to a conventional view of a list of all studies, the design concept strives to combine overlapped search results by intervention types (intervention overview), because many clinical trials investigate the same intervention but in different medical centers or locations.

Intervention overview

Result content

[Treatment search | myTomorrows](#)

Saved results [Search report](#) [Subscription](#) [Sign up](#) [Sign in](#)

[Home > COVID-19](#)

COVID-19 

Suggested terms for 'COVID-19'

2 treatment options found for "COVID-19"

All studies [Intervention overview](#)

1 Remdesivir 

Remdesivir is a broad-spectrum antiviral medication developed by the biopharmaceutical company Gilead Sciences. As of 2020, remdesivir is being tested as a specific treatment for COVID-19, and has been authorized for emergency use in the U.S. and approved for use in Japan for people with severe symptoms.

2 7 Studies (6 Clinical trials CT 1 Expanded Access Programs EAP) 

3 GILEAD Creating Possible 

Hydroxychloroquine (Plaquenil) 

Hydroxychloroquine, sold under the brand name Plaquenil among others, is a medication used to prevent and treat malaria in areas where malaria remains sensitive to chloroquine. Other uses include treatment of rheumatoid arthritis, lupus, and porphyria cutanea tarda. It is taken by mouth.

20 Studies (20 Clinical trials CT) 

Producer 

Contact us 

Intervention Overview page 

1. Intervention title and a short description
2. All linked studies (CT/EAP) to the intervention
3. The producer (e.g. pharmaceutical company) of the intervention

The intervention overview page curates all treatment options by binding all clinical studies that investigate the same intervention and other additional information, for instance, the pharmaceutical companies behind. Besides, the studies (2) will lead users to the search results with

the new search query (e.g. Remdesivir) standing for the intervention.

Note that the intervention options in the image above do not represent the reality that there are many more investigational treatment options for COVID-19, and it is a mere concept.

1 [Home](#) > COVID-19 > [Study to Evaluate...](#)

2 [Study record](#) [Study design](#) [Related study result](#)

3 Description

The primary objective of this study is to evaluate the efficacy of 2 remdesivir (RDV) regimens with respect to clinical status assessed by a 7-point ordinal scale on Day 14.

4 Process

- 01 Search
- 02 Review**
- 03 Discuss

5

- Description
- Eligibility Criteria
- Locations
- Additional information

6 [Back to result list](#)

Result Content page

Contact us

Discuss for access

Inclusion sex All

Inclusion age 12 Years and older (Child, Adult, Older Adult)

Inclusion criteria

- Willing and able to provide written informed consent, or with a legal representative who can provide informed consent, or enrolled under International Conference on Harmonization (ICH) E6(R2) 4.8.15 emergency use provisions as deemed necessary by the investigator (age ≥ 18), or willing and able to provide assent (age ≥ 12 to < 18 , where locally and nationally approved) prior to performing study procedures

Aged ≥ 18 years (at all sites), or aged ≥ 12 and < 18 years of age weighing ≥ 40 kg (where permitted according to local law and

Contact us

- Primary study information with actions
- Study content by category
- Study content
- The process graph
- Dynamic content directory
- Return action to search result page

The result content page comprises a CT/EAP record from a clinical register, actions (top right corner), and guidance information for leading to the contact form page. The textual content is put into three sub-pages of Primary record, Study design, and Related study result to improve the

readability and efficiency. It is also assisted with a dynamic content directory (5) that gives a quick overview of the information structure and a shortcut to access the desired content.

5.5.4 Three synthesized feedback concepts

The iterated baseline search serves as the foundation for accommodating relevance feedback collection concepts. On top of it, three relevance feedback collection concepts were synthesized, based on the insights gained through the speed dates of the nine feedback ideas (p.94-96), ideas generated from brainstorming sessions (p.106-107), and the revision of the theories (p.41-45).

Instead of creating one concrete relevance feedback collection concept, it is necessary to keep the abundance to answer the research questions (p.25) that require a comparison of different factors. Also, the results from the previous user tests could not represent the target users (healthcare professionals).

All three concepts try to maximize the simplicity in interactions and experiment with different motivators. The first concept employs personal utility, and the second tries to trigger users' altruism by making the feedback collection request more direct and apparent. The last one attempts to engage users through enjoyment.

Concept 1 - Pre-screener

The first concept, Pre-screener (on the next page), aims at engaging users through personal utility. It is a variation of the feedback idea eight (p.96), but the difference is that it reduces the interaction complexity and enhances its functionality. Figure 61 illustrates the simplified flow of this concept.

By allowing users to mark each search result with their interest (a primary type of relevance) in the search result page, users could manage results, be recommended with similar results, and hide irrelevant ones from the search result list. There are three symbols (representing interested, not interested, and not sure) appearing on the right side of the search result by default. The mark will be highlighted when clicked, and the search result will be added to the corresponding folder that appears on the top of the search result list. On hovering on the cross (not interested), options of reasons tags will show up. Once marked, the result will grey out in the search result page.

Each folder accommodates marked results but differs in the functionalities. The interested folder will have recommendations based on these marked as interested. In contrast, the not interested folder will allow users to hide them from the search result page and tagging reasons for irrelevance.

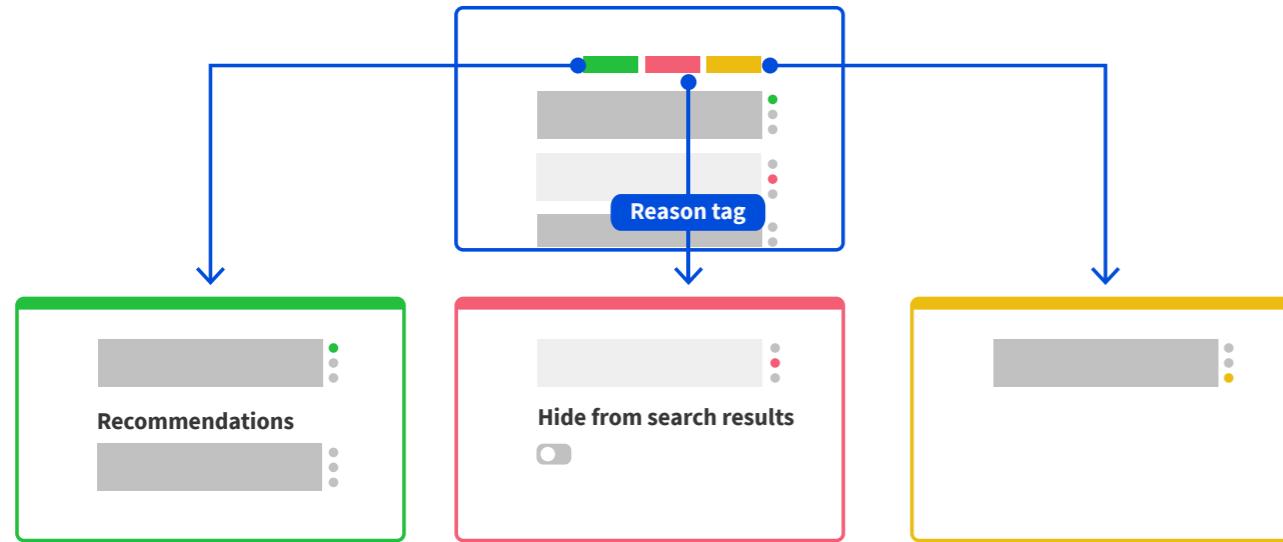


Figure 61. Illustrated logic flow of Feedback concept 1 (Pre-screener, Drawn by author)

Concept 1 - Pre-screener

Try it out (Prototype)

This screenshot shows the 'Treatment search | myTomorrows' prototype interface. At the top, there's a search bar with 'COVID-19' and a magnifying glass icon. Below it, a message says '400 results found for "COVID-19"'. The main area has tabs for 'All studies' and 'Intervention overview'. A 'Prescreening tool' section includes buttons for 'Interested (1)', 'Not interested (1)', and 'Not sure (1)'. To the right, a list of four study entries is shown, each with a checkbox, status, phase, and title. The first entry is 'CT Recruiting' for 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'. The fourth entry is 'CT Recruiting' for 'Valsartan for Prevention of Acute Respiratory Distress Syndrome in Hospitalized Patients With SARS-CoV-2 (COVID-19) Infection Disease'. A modal window titled 'Mark as Not Interested' is open over the list, listing reasons: 'System failure', 'Won't help my patient', 'Other', and 'Create my reason tag'. A 'Contact us' button is also visible in the modal.

This screenshot shows the same prototype interface after marking the first study as 'Interested'. The 'Prescreening tool' now shows 'Interested (2)', 'Not interested (1)', and 'Not sure (1)'. The study list shows the first study as 'Interested' (indicated by a green checkmark icon). The 'Recommendations' section of the modal window is visible, stating '2 Recommendations' and 'Recommendations are based on those you have marked Interested, and it will adapt to your choices.' The 'Not interested' section of the modal is also visible.

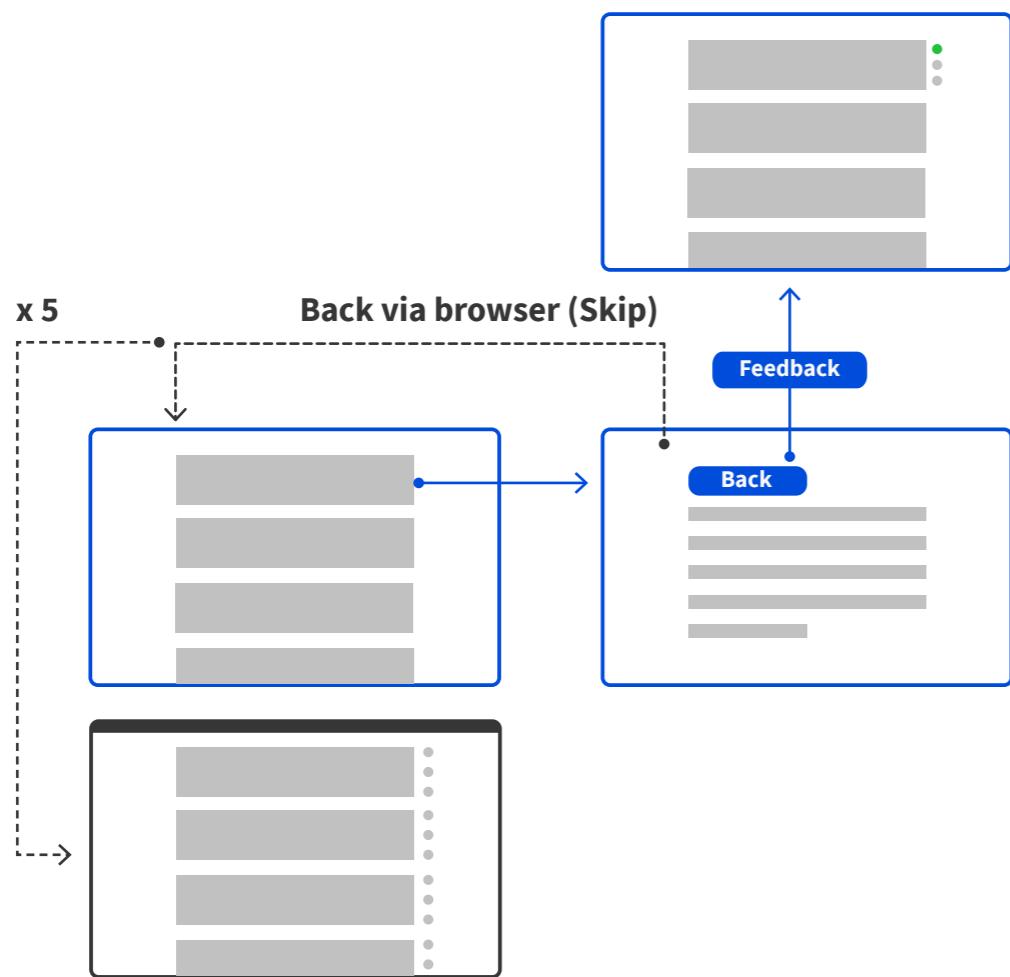


Figure 62. Illustrated logic flow of Feedback concept 2 (Reminder, Drawn by author)

Concept 2 - Reminder

This concept, Reminder (on the next page), is a combination of feedback idea one and six (p.93), as well as the inspiration from the brainstorming session (p.104). It inherits the motive of altruism by prompting users with a relevance feedback request that replaces the 'back to search result' button when users are about to return. The request follows the standardized mechanism (p.90), and when any options are clicked, it works as the same function as a back button.

If users use alternatives (e.g. trackpad) or keep skipping feedback requests (Figure 62), a reminding message will show up underneath the navigation bar, and relevance marks will appear on the right to the reviewed search results (after reviewing five results but not giving any feedback). Such a message stems from the feedback idea seven (p.94), but the key difference is that relevance feedback could be given directly through the marks on the right of the search results. (similar to concept 1).

It may sound against any design principles and intuitively wrong, but it was designed to replace the back button on purpose, rather than putting a feedback request close to the back button. In this way, it nudges users to provide feedback

The screenshot shows the 'Treatment search | myTomorrows' interface. At the top, a blue header reads 'Concept 2 - Reminder' and 'Try it out (Prototype)'. Below the header, the search results for 'COVID-19' are displayed. One study is highlighted: 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'. The study details include: Type: Clinical trial; Status: Recruiting; Phase: 3; Start date: March 6, 2020; End date (est.): May 2020. The 'Study record' tab is selected. On the left, a sidebar shows the process: 01 Search (grey dot), 02 Review (dark blue dot, currently active), and 03 Discuss (light blue dot). A feedback prompt at the bottom left says: 'Is this study useful?' with three options: 'Yes' (green checkmark), 'No' (red X), and 'Not sure' (yellow question mark). It also states: 'It will turn back to the search result page automatically once feedback is provided'.



The screenshot shows the 'Treatment search | myTomorrows' interface with a yellow header bar indicating 'You have reviewed 5 results and given 1 feedback, leave feedback on all reviewed results by marking the symbols. This will help improve the search experience'. Below the header, the search results for 'COVID-19' are displayed. The results table shows 400 results found for 'COVID-19'. The first result is the same study as in the previous screenshot: 'Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)'. The results table includes columns for 'All studies', 'Intervention overview', and filters like 'Type', 'Status', 'Phase', 'Study title', 'Intervention', 'Save', and 'More'. The results are listed with columns for 'ID', 'Status', 'Drug', and 'Title'. The first result has a green checkmark icon and the drug 'Remdesivir' listed under 'Drug'.

Concept 3 - Hearty

**Try it out
(Prototype)**

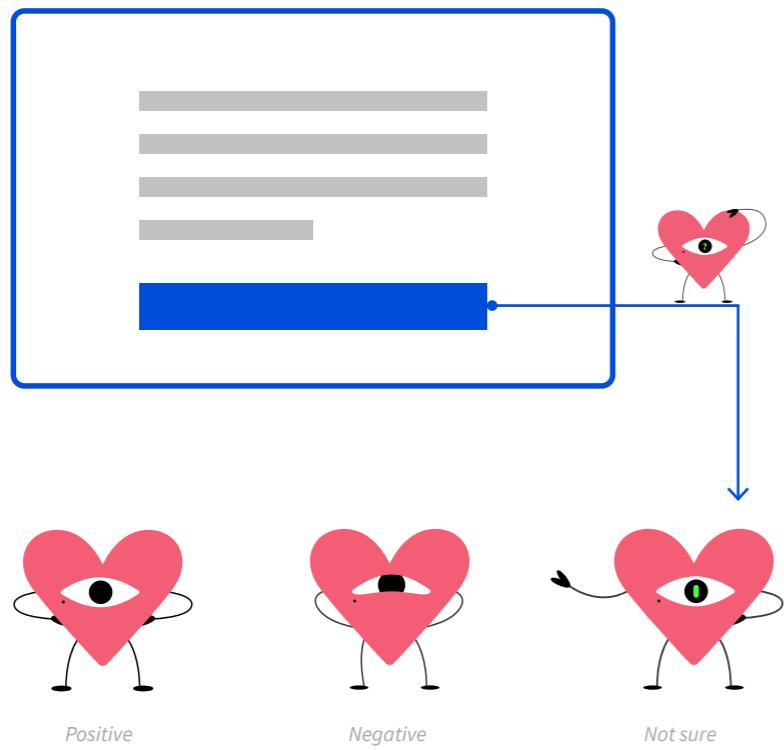


Figure 63. Illustrated logic flow of Feedback concept 3 (Hearty, Drawn by author)

Concept 3 - Hearty

The last concept, Hearty (on the next page), is a continuation of feedback idea nine (p.94). It aims to trigger internal motivation (personal enjoyment) and borrows motivational affordances from the Gamification theory. For instance, badges, levels, and progress (Hamari et al., 2014). Unlike the original idea, the avatar in this concept tries to imply a health context with the design of a colorful, warm, and heart-shaped character named Hearty.

Hearty appears at the bottom of each search result with the standardized feedback request. When users give feedback by answering the question about relevance, it responds (Figure 63) based on the relevance selections: a happy face to positive options, a sad face to negative options, and an encouraging face for not sure. Besides, Hearty will grow along with the amount of feedback (visualized with a progress indicator) provided by users. Moreover, users will receive badges as achievements to encourage them to provide more feedback (e.g. first feedback achievement).

The screenshot shows a web application interface for 'Treatment search | myTomorrows'. At the top, there's a navigation bar with links for 'Saved results', 'Search report', 'Subscription', 'Sign up', and 'Sign in'. Below the navigation, there's a section for 'Locations' listing countries like United Kingdom, Singapore, Spain, Sweden, United States, Taiwan, Netherlands, and Japan. Under 'Additional information', it shows 'Trial ID: NCT03469609' and 'Sponsors: Hannover Medical School, German Research Foundation'. The main content area features a Hearty character asking 'Is this study interesting for you?' with options 'Yes', 'No', and 'Not sure'. A message says 'You haven't provided any feedback yet.' Below this, a red 'X' icon and the text 'Sorry to hear that it's not what you are looking for.' are displayed, with a link 'Change the feedback'. A progress bar indicates 'Provide 4 more feedback to receive your next badge' with '1/5' completed. To the right, a large Hearty character with a progress bar labeled 'View all achievements' is shown. At the bottom, a progress bar for 'First feedback' is at '1/5', and a row of five achievement badges is visible: 'First feedback' (blue), '5 in a row' (blue), 'Feedback master' (purple), and '100 feedback' (orange).

Key takeaways chapter 05

5.1 Creative session for design ideation

The results of the brainstorming session with four peers imply that peers perceive trust more on the information level through its usage and recency. Surprisingly, peers didn't expect much guidance in search. Instead, they would reflect on their competence. All participants felt odd about providing explicit relevance feedback in search partially because of the lack of experience.

The ideas indicate that perceived trust could be increased by displaying collective relevance, increasing system explanation, and enhancing search functionality. In terms of system guidance, it could be enhanced by making the result matching process more interactive, recommending search queries, linking external resources, improving reading efficiency, and personalizing search processes. Ultimately, the ways to engage users to provide feedback is to combine feedback provision with functions, use simple questions, and affiliated feedback.

5.2 Baseline search construction

The information architecture adapts from the original one following the design requirements with twelve significant changes. In terms of the user journey, the concept starts with a filling out of a one-time questionnaire to classify different user groups. HCP users or other groups with sufficient medical knowledge to make relevance judgments will be led to a regular search interface with enhanced relevance feedback collection. Whereas, patient users will be directed to a question-based interface. The main focus of the project is on collecting HCPs' relevance feedback on search results, thus, the conceptual baseline interface incorporates five major feature user flows (core user flow of search, Subscription, Search report, myFolders, and Suggest study results) following the design requirements.

5.3 Feedback ideas generation

Inspired by previous research, nine feedback collection ideas were generated and built in the baseline search based on the same feedback collection mechanism that HCPs can opt-in to provide feedback. Once they are in the flow of feedback provision, there will be three options (positive, negative, uncertain) to the question of inquiring relevance judgment. At the end is a thanks message, and HCPs will be allowed to change given relevance feedback. All the nine feedback options belong to three stages (respectively pre-screening feedback, in situ feedback, and post hoc feedback) and incorporate different motivators of utility, altruism, enjoyment, and challenge.

5.4 Usability testing and ideas speed date

The usability testing and ideas speed date with myTomorrows employees with medical background and peer students suggest that the baseline concept slightly outperforms the original myTomorrows site for the core flow of search and the overall impression. Despite the differences among all participants, all internal HCPs preferred the baseline concept design as it offers more functionality and could potentially increase their working efficiency. Besides, the prioritization of all new features indicates that Search report is the most wanted feature, followed by myFolders and Suggest related results. However, Subscription has the lowest priority.

Among all nine feedback ideas, the first idea (Bottom pop-up) is the most liked idea with three favorite votes and one most disliked vote. In comparison, the seventh idea (Task session) gathers the highest amount of most disliked votes (three without any favorite votes). The rest are not outstanding in terms of participants' preferences, but the last two ideas remain controversial. Through the heuristics of participants, it is learned that eight points matter: the value of providing

feedback, the visibility of the feedback request, timing of showing up, the embodiment of the feedback request, perceived effort, control, tone, and context fitness of the feedback request.

5.5 Feedback enhanced search conceptualization

The results from another round of brainstorming session on feedback collection show the pattern that feedback collection has a strong relation to the stages within a search process, which is in line with the three categories (Pre-screening feedback, In situ feedback, and Post hoc feedback) of feedback ideas.

Based on the test results from previous research (section 5.4), three relevance feedback concepts were synthesized in the iterated baseline search: Pre-screener, Reminder, and Hearty. The first concept (Pre-screener) incorporates utility as the motivator. It collects relevance feedback on the search engine result page. The second and the last concepts belong to the same category in situ feedback as they both try to engage users to provide feedback after reading the content in the search content page. The only difference is on the underlying motivators that the former incorporates altruism and minimizes the ability required from users. The latter, in comparison, uses enjoyment as the motivator to engage users.

06

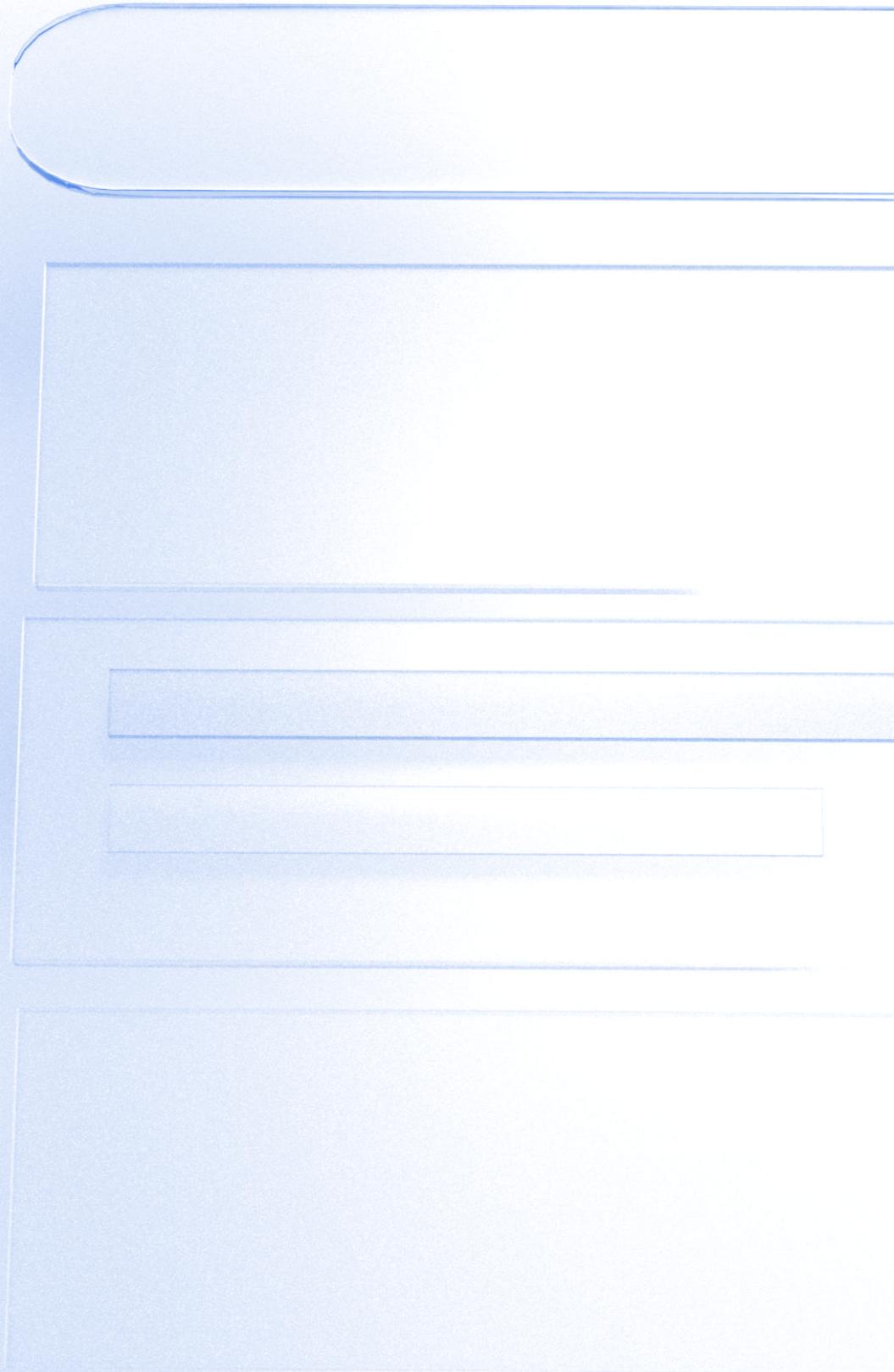
Concept evaluation and generalization



Chapter six brings the Search enhanced with three relevance feedback collection concept to nine end-users (healthcare professionals) for evaluation. The insights gained from the sessions were generalized into a guide for designing explicit feedback collection comprising a four-stage process and ten recommendations.

Chapter overview

- 6.1 Concept evaluation with target users
- 6.2 A guide for designing explicit feedback collection



6.1 Concept evaluation with target users

The baseline Search with three built-in relevance feedback collection concepts was used as the research material to probe the target users (external HCPs). There are two objectives 1) elicit external HCPs perspectives on the usefulness of the overall search, and 2) understand which feedback collection concept would better engage them in providing feedback.

6.1.1 Methods

The user evaluation session involved nine external HCP participants* from three countries (Netherlands=6, China=2, Brazil=1), recruited via personal connections, and the support from myTomorrows. Nearly all (5) Dutch HCPs were from Erasmus MC, and more than half (6) had no more than five years of experience in clinical practice.

Note: four internal HCPs participated in the evaluation, but it was excluded because of a strong bias revealed. Data from internal HCPs were documented but annotated for comparison purposes.

All test sessions (each ranges from 30-60 minutes) were performed individually with HCPs (Figure 67) via video conference software (Zoom and Skype). They all followed the same test procedure that consists of four sections illustrated as below): briefing, pre-test questionnaire, test, and post-test questionnaire and interview.

The briefing section introduced the project context as well as the objectives of the project. By presenting an example of feedback collection, the pre-test questionnaire aimed to gather HCPs' attitudes towards feedback provision and help them recall their previous experience of feedback provision. Participants were asked to think

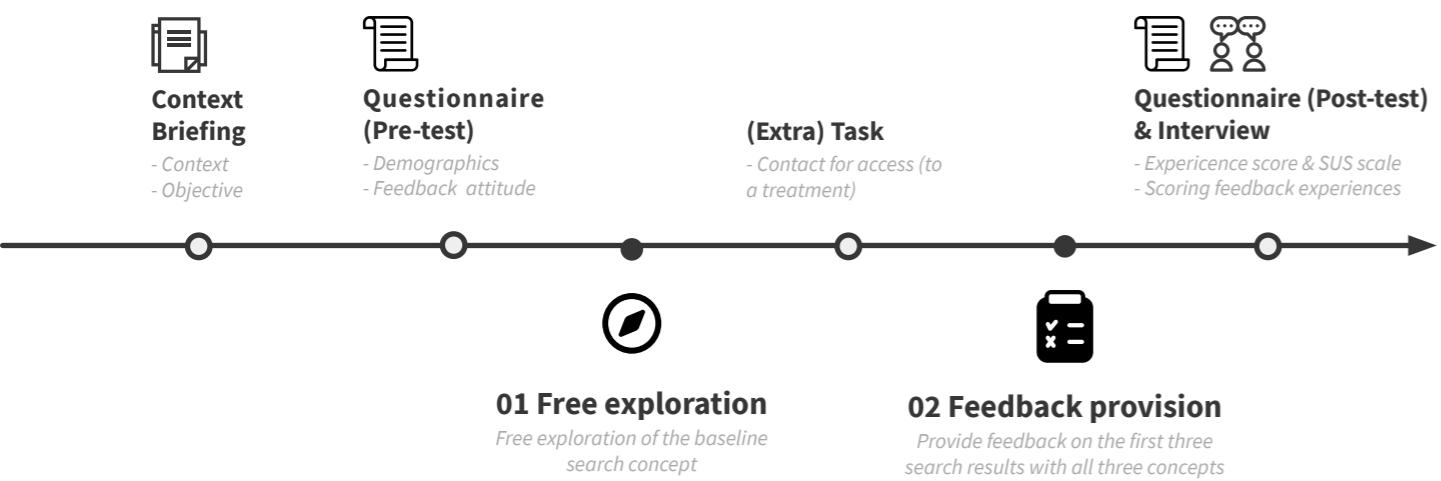
aloud during the test section that included two tasks: 1) free exploration of the baseline search and 2) feedback provision in three concepts (p.116-121). The evaluation session ended with a questionnaire and a semi-structured interview on their experiences, rationale and explanation of their choices filled in the questionnaire.

Each evaluation session was screen recorded, starting from the test section with informed consent from participants.

Test section

The test section took approximately 20 minutes, which was allocated to each task evenly. In the former task, participants were asked to freely explore the baseline Search concept. It could provide a comparison of the search experience with and without feedback collection. An extra subtask was assigned to those who did not experience the flow of gaining access to a study.

In the other task, three relevance feedback collection concepts were presented to HCPs following the same sequence for performing the same task of providing relevance feedback on the first three search results.



6.1.2 Test results

The raw data of user evaluation sessions included answers from the questionnaire and the transcribed session records. In general, the analyzed results comprises three parts: 1) the experience (SUS) score, 2) impression with the baseline Search, and 3) insights on three relevance feedback collection concepts.

1) Overall experience and System usability scale (SUS) scores

The baseline Search experience was quite decent as the mean score of the overall experience rated 4.11 (out of 5, Std=0.74, Figure 65). Besides, the ten SUS questions for evaluating the perceived usability resulted in an overall score of 75.31 (out

of 100, Std=9.63), and individual scores ranging from 90 (highest) to 60 (lowest). However, the SUS result excluded the data from HCP 8 because she expressed her concern that her input may pollute the dataset, so the individual score of each SUS question was either 1 or 5.

Figure 66 visualizes both the overall and individual SUS scores in a mixed graph that combines grade ranking of SUS score graph (Banger et al., 2009) and the percentile rank graph (Sauro, 2011). The former associates the SUS scores with adjectives, while the latter links SUS scores to the distribution of all systems that have tested with SUS. The results indicate that the baseline Search was perceived beyond good, and the average score is higher than (approximately) 70 percent of all products that were tested with SUS and could be graded as B.

How was your overall experience with the treatment search?

9 responses

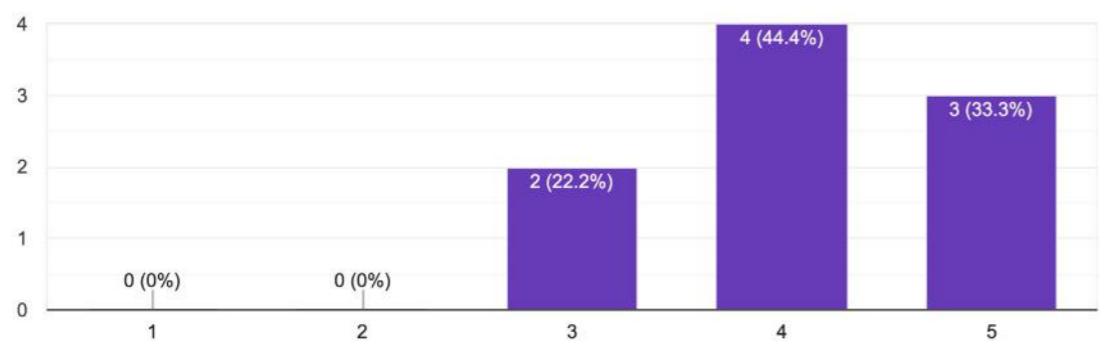


Figure 65. Overall experience score (Questionnaire result)

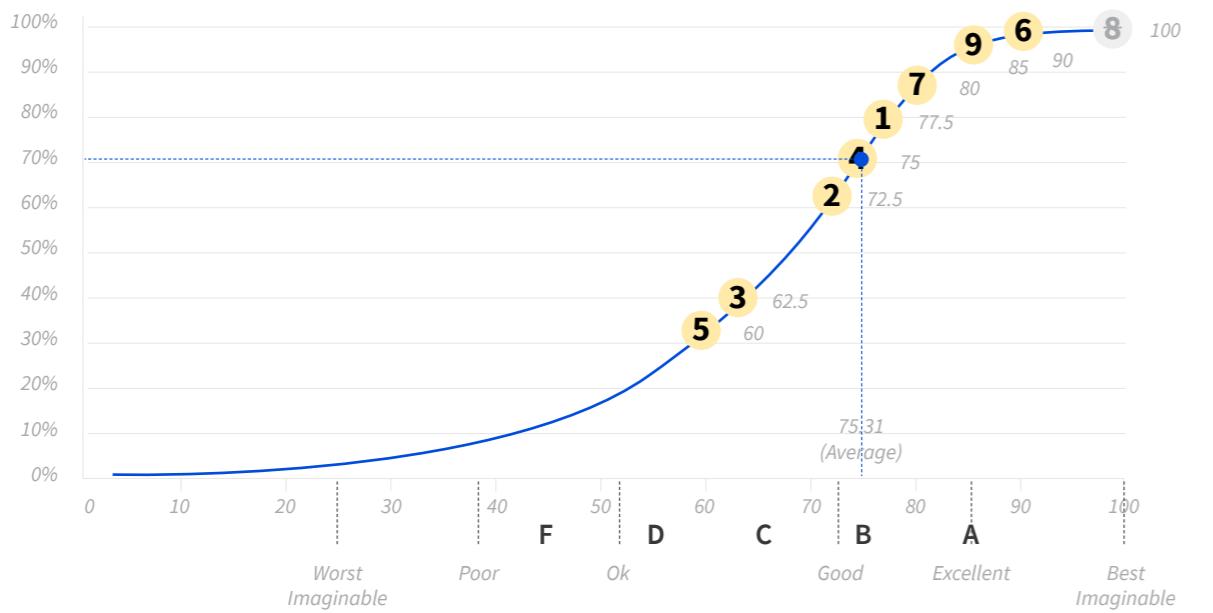


Figure 66. A mixed SUS score graph combining percentile rank and grade ranking (image source: Banger, Sauro, adapted by author)

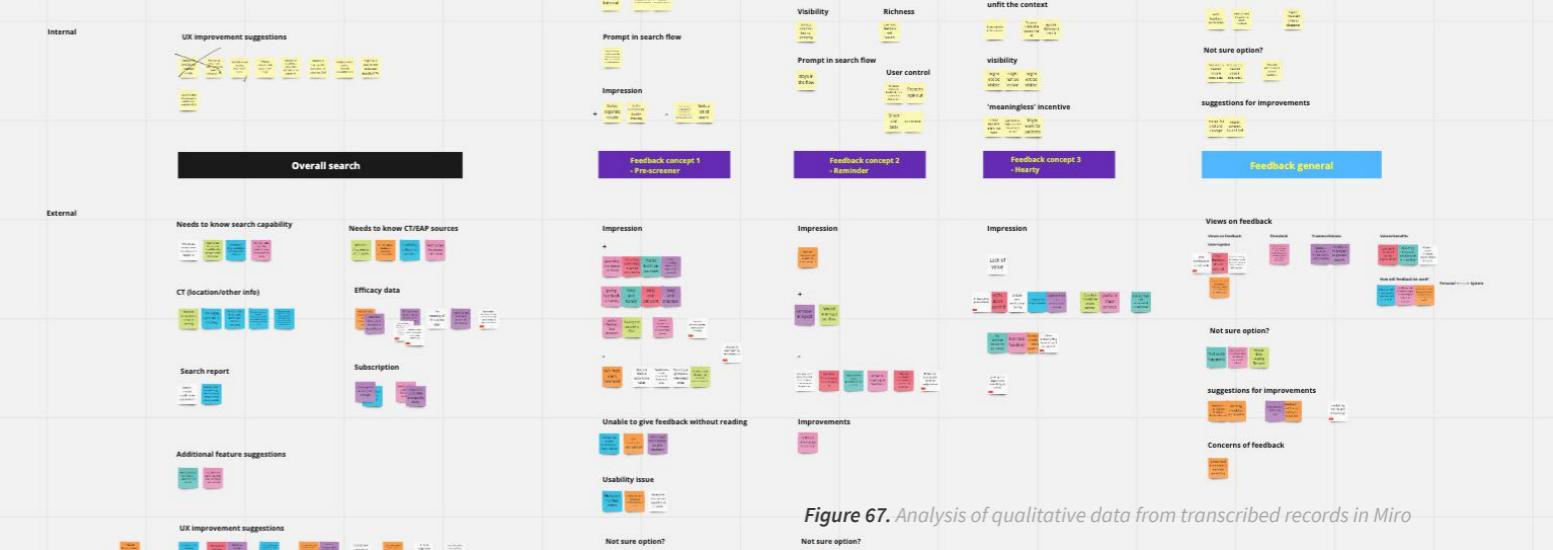


Figure 67. Analysis of qualitative data from transcribed records in Miro

UX complaints	<ul style="list-style-type: none"> Confused of the repetitive buttons 'discuss' and 'discuss for access' (content page) The meaning of 75% survival rate is ambiguous (related study results) Expected to gain access to an intervention directly from the intervention overview (Intervention overview) Replace doctor to health care professional (website) The information required in contact form is not clear (e.g. patient's name or HCP's name) or missing (e.g. HCP's institution) (Contact form) Adverse complaints may give HCPs a false impression that the search engine provider is affiliated to a pharmaceutical company (adverse complaint, footer) How mytomorrows service works should be better communicated to avoid bias or promotion (general)
Facts	<ul style="list-style-type: none"> Eligibility of CTs is the prior concern when finding treatment options for patients Downloading potential CT treatment is most common way to present to patients
Additional feature suggestions	<ul style="list-style-type: none"> Supports results export Wants to track a single study updates (subscription) The naming may need reconsideration to mitigate the confusions (search report) Includes more patient profile filters (e.g. ethnicity, comorbidity) Customizable efficacy data visualization (internal HCP) Include email and phone number in live chat (internal HCP) Specific location for studies (internal HCP) Replace N/A with EAP to increase credibility (internal HCP) Expects a way to see reviewed results (internal HCP)

Table 14. Overview of external HCPs' impressions and experience

2) HCPs' impression on the baseline Search

Transcriptions were analyzed (Figure 70) following the Grounded theory. In general, external HCPs had a positive impression on the baseline search, and they perceived it as user-friendly (more than Clinicaltrials.gov). What attracted them the most

was the integration of multiple databases and the efficacy data overview. However, HCPs did not expect to have rich functionalities (Subscription, myFolders, Search report) in a search engine, but having them would be a plus. Tables 14 shows an overview of external HCPs' impression and experience with the baseline search. Besides, there are seven key points extracted as shown in the next page.

Search capability is the core

The search capability was the essential aspect external HCPs were looking forward to from a search engine. They would typically use different search queries to examine the capability, and the criteria depend mainly on whether the information retrieved is of good quality, highly relevant, or recent. However, this research couldn't allow it.

HCP 1, "I would use more terms than COVID-19, if I was searching a specific thing for my patient."

HCP 5, "I think it's easy to use. But how do we know as a healthcare worker that the database is totally up to date? That's to make sure before I want to use it."

Data sources matter

Next to the search capability, the data source was one crucial aspect external HCPs care about, as is linked to the credibility of the information. HCPs wanted to make sure the information retrieved in myTomorrows treatment Search is authorized and unbiased. They have a strong urge to know the information source immediately (similar to medical articles from different journals could give a hint to HCPs), indicating that the source needs to be better communicated and visually more prominent.

HCP 4, "Just maybe you could just mention in the tech software that it searches multiple databases."

HCP 5, "And can you see the journal it's in the SERP (source), or only when you click on it, so which databases were connected to this one?"

HCP 9, "I'm not quite sure where this information came from. Also....it seems to be sponsored, so I'm a bit hesitant to always read about it. I think it's mainly because of the sponsor there. This is also the kind of data you can see in advertisements in a medical journal from any kind. I mean, I think it's mainly the logo, it gives me doubts about how independent this data is and where it comes from."

More abundant CT information is preferred

For external HCPs, they would want to know as detailed information of a study as possible in order to make any well-considered clinical decisions. The Search design concept provided primary information to them, but the information provided

in CT was incomplete (compared to its source). For instance, the specific running locations of a CT and its sponsor(s). Such information is vital for external HCPs as they could make a primary judgment of trial quality or potential biases.

In addition to the aforementioned missing information of a CT, the Search lacked information or instructions (e.g. voluntary and safety disclaimers) aimed at communicating to patients, since they are the ones who make the final decision to participate in an investigational treatment or not.

HCP 1, "I would want to know where the trials are running, at which medical center?"

HCP 2, "There is another question that, normally clinical trials have safety description, informed consent, disclaimer, and etc. it doesn't include that information, right? If you're gonna present a CT to a patient, because it depends on them to make the final decision, but...it will not be a problem if the company that is arranging everything can provide this information when it is about to access. But it is still not clear where I can get that information."

Intervention efficacy data needs more clear communication

External HCPs were pleased to see an overview of previous study results. However, they were curious how and where all these data were from and curated because the credibility of information is the prior concern they have on CT/EAP. Furthermore, some participants had the impression that the data overview shown was real-time data. Thus, they pointed it out the infeasibility of displaying immediate efficacy data of an intervention under investigation.

HCP 4, "But in general, by the time you have this available (published study results), then studies will no longer be enrolling patients."

HCP 7, "Real-time data in clinical research, sharing. It never happens. Because as soon as you share your data, someone else can publish probably. It happened quite often during the COVID-19. I'm really curious about how you would be able to get real-time data on studies which are still active because if you're active as a study, you almost don't know, it's very strange to have like, real-time data."

Discussion brings convenience but is not the primary path for access

The core flow within the baseline Search was clear to all participants as no one struggled in finding the entry point for accessing a study or intervention. As discovered in the exploratory phase (Chapter 3), external HCPs would prefer directly contacting the medical center that runs the Clinical Trial in which they are interested. However, with the design concept, some HCPs expressed that contacting a Clinical Trial through the site (via live chat or filling out the contact form) would be more convenient than looking for the contact information (of a medical center and the investigators) and waiting for responses.

HCP 6, "Okay, so that's also interesting though, you would be able to contact the people doing a study by your website. So that's interesting. It would make it easy to have to look for emails or phone numbers."

HCP 7, "I wouldn't expect the search engine to have discussion options. So I think that's an extra. Yeah, would be interesting to me."

HCP 9, "Um, well, I might have some, you know, some questions directly related to the study, or maybe I would ask them to get me in touch with researchers, although that's something I would do through other routes. No, I would contact the principal investigator directly."

Subscription could be more clear and powerful

Some HCPs thought that subscription did not communicate clearly about its functionality. In contrast, two HCPs understood the purpose well and expected a more powerful subscription functionality. They expressed the potential need of subscribing to a specific study or for a specific patient. So that they could keep tracking a study's updates because it happens that the study record needs updating or refining even when it is available in clinical registers.

HCP 5, "Sometimes when I find a clinical trial that's going to start in one week, but a week later, they update the protocol. Do I get this email of the updated protocol, or do I have to find it out by myself? I use clinical trials for one of my studies, and we have updated the protocol for a few times already. So for people who want to know more about the newest protocol."

HCP 7, "Is it possible to save certain studies when you haven't subscribed to the page assigned to the search

engine? And then when you subscribe, it's possible to save certain studies into may be to follow through time or whatever. So subscribe to a certain disease. This is a COVID-19 search, but it's also with the patient data. So actually, you don't only subscribe to COVID-19 but also to COVID-19 for a man or a woman who is 52. For this patient, I have seen this in this study, and you would actually want to have like subscriptions per patient."

Search report is ambiguous

Not all HCPs noticed Search report in the menu, and for those who discovered and looked into it, they expressed their confusion that it was not clear what Search report is about and how it works. Moreover, one HCP assumed it contains published study results. This result indicates that Search report, as the main conversion channel, needs strong and informative communication through various channels.

HCP 1, "I click the search report, and I'm here. Yeah, apparently, the amount of trials has been reduced quite a bit. Certainly not really sure what it is about, okay. tells me to enter a code that I'm supposed to receive but not really sure."

HCP 2, "Why is the search report thingy here? I cannot figure out what it is designed for, maybe if there are published study results in 'search report'..."

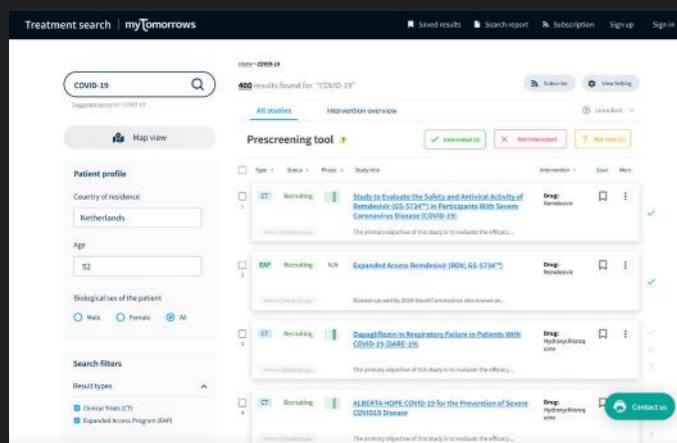
3) Comparing three concepts

The questionnaire results on external HCPs' experience with all three concepts revealed that, the first concept (Pre-screener) outperforms the other two concepts and is more preferred in general. However, preference does not imply that it is more motivating or needed. Instead, the second concept slightly exceeded the first one on motivation, and the third concept did not fall far behind.

Concept one is preferred but the second is slightly more motivating

Overall, the first concept had the highest experience score (Figure 71) of an average of 3.78 (out of 5, Std=1.23), but the distribution went two extremes. The second concept (reminder) received an average of 3.56 (Std=0.83), and the last concept had the lowest average score (3.22, Std=1.03).

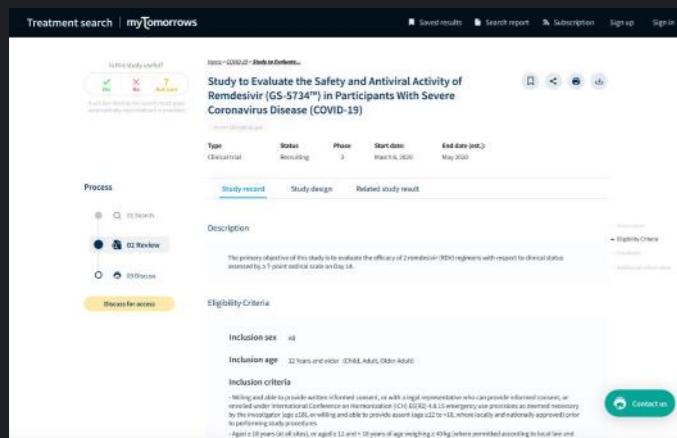
Concept 1 - Pre-screener



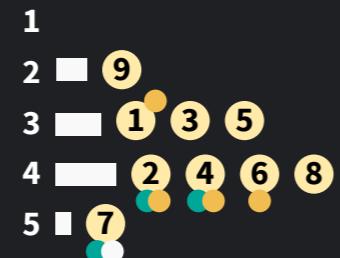
**Average score: 3.78/5, Std=1.23
Fav=6, Most motivat=3**



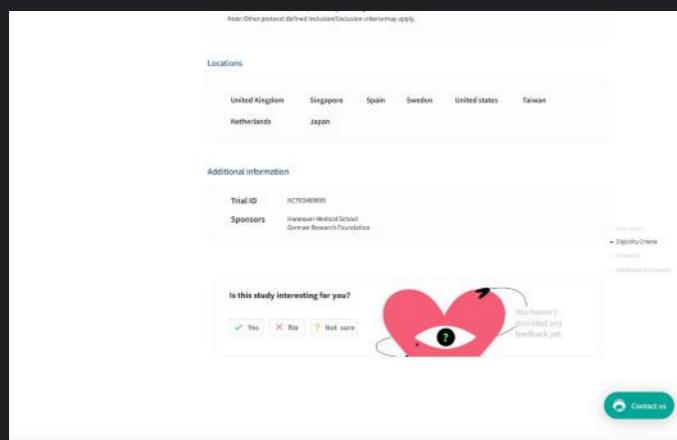
Concept 1- Reminder



**Average score: 3.56/5, Std=0.83
Fav=3, Most motivat=4**



Concept 3 - Hearty



**Average score: 3.22/5, Std=1.03
Fav=0, Most motivat=2**



● Favorite ● Most motivating ● Prefer having ● n external HCP participants with coded number

Figure 68. Scores and votes on each concept

Two-thirds (6) of the participants selected the first concept as their favorite one, and three picked the second concept. However, no one preferred the third concept (Hearty).

Surprisingly, votes for concept (4) two as the most motivating concept surpassed the first concept by one vote. It was because the second one is more direct and natural to interact with. Nevertheless, over half (5) of the participants were self-consistent that they regarded the favorite concept as the most motivating one as well. Conversely, four external HCPs did not, and two of whom voted the third concept as the most motivating concept.

**The results from internal HCPs revealed a more unified outcome that most (3 out of 4) both preferred and regarded the second concept as most motivating. Only one chose the first as the favorite and the third as most motivating. It may be because the value of providing feedback has been internalized, and the more direct, the more they would prefer.*

Preference does not represent necessity

Participants were asked whether they would prefer having the favorite concept in the search. Only four would prefer having the favorite concept to be built in the baseline search, while the rest felt neutral. Among the four, three selected concept one as the favorite, which accounts for only half

of the participants who preferred the first concept over the other two. In comparison, only one out of the four was for the second concept, and the number is merely one-third of all favorite votes for the second concept.

The results indicate that the preference may not be equal to wanting it. Besides, using utility as the motivator would be more appealing to users.

Concept one 'wins' the competition, but it could learn from the other two

Table 15-17 listed the pros, cons, and general feedback on all three concepts. Overall, the first concept outperforms the other two concepts as it gathered most favorites and similar votes for the most motivating concept. More importantly, three external HCPs would want to have it built in the Search. However, the first concept has the drawback that it collects the primary relevance judgment made before reviewing the content. Thus, it could lead to shortsighted feedback, or users need to remark relevance afterward. Nevertheless, combining the second concept with the first could potentially resolve the issue. Moreover, the first concept could borrow the rewarding (levels and badges) mechanism from concept three but in a more serious manner.

Concept 1 - Pre-screener	+ Pros	<ul style="list-style-type: none"> It helps build a sorted overview of all search results, especially valuable for a large number of document reviews
	- Cons	<ul style="list-style-type: none"> (Usability) Relevance marks are not primarily too visible, and it may collide with the checkboxes (Usability) Relevance folders do not seem clickable The feedback could be superficial, shortsighted, or the snippet information is not sufficient to make a judgment without reading the content Not all participants felt it useful, especially with the negative marks because some HCPs thought they could filter irrelevant ones out in the head by merely ignoring. Besides, saving could be a replacement for marking a result as relevant The not sure option is controversial because in whatever cases, results marked as interested or not sure need further reviewing
	General	<ul style="list-style-type: none"> Users want to know the recommendation criteria (in the interested folder), to make sure they are not promoted

Table 15. Synthesized experiences external HCPs had on the first concept

Concept 2 - reminder	+ Pros	<ul style="list-style-type: none"> It prompts at an appropriate timing Users will be attracted, and it does not interrupt reading More detailed relevance feedback can be collected than the first concept as it collects feedback after reviewing the content
	- Cons	<ul style="list-style-type: none"> Users felt forced to provide feedback and they cannot opt out When in a rush, random or invalid feedback could be collected
	General	<ul style="list-style-type: none"> Users do not perceive much value compared to the first one, and it does not seem to be of help for improving the search capability As negative feedback accumulates, users would be annoyed. (There is no exact reason mentioned, but it can be linked to their dissatisfaction with the search engine's performance.)

Table 16. Synthesized experiences external HCPs had on concept two

Concept 3 - Hearty	+ Pros	<ul style="list-style-type: none"> The badges could trigger intrinsic motivation (HCP 9)
	- Cons	<ul style="list-style-type: none"> It does not fit much in such a serious context and was perceived as childish and not serious enough It was perceived as an ad banner (HCP 8) because of its colorfulness and playfulness design Due to the position, the feedback request may be missed when the relevance judgment could be made without scrolling down to the very bottom
	General	<ul style="list-style-type: none"> Having an avatar is acceptable, but its embodiment needs to be more serious Some thought feedback request should be direct to the point The cuteness of the avatar could be more for patients or internal HCPs, but not external HCPs Users do not perceive much value compared to the first one, and it does not seem to be of help for improving the search capability (the same as the second)

Table 17. Synthesized experiences external HCPs had on concept three

6.1.3 Insights on relevance feedback

Through heuristics extracted seven major insights on external HCPs' attitude and experience towards relevance feedback provision. Some insights share in common with previous findings (p.103-104) from peers and internal HCPs. Nevertheless, what essentially distinguishes the insights below is the external HCPs' need for understanding

feedback data usage and the ultimate impact on the seeking information.

1) Providing feedback does not make external HCPs feel negative

Through the questionnaire (Appendices, p.195-206), the majority of external HCPs (7 out of 9) would provide feedback in some cases. However, two HCPs expressed that they never provide feedback in the given (i.e. Google help) or similar

situations. The reasons why HCPs chose not always providing feedback may vary from person to person. Nevertheless, there is the consensus that external HCPs do not feel negative about providing feedback, as over half (6) external HCPs felt neutral towards feedback provision, and three felt positive.

2) Feedback collection should prompt gently at the right time and be optional

Like the previous finding that people generally think that providing feedback is an extra work, external HCPs are not exceptional. They regarded feedback provision as a side task deviates from their goals of looking for relevant information in a search system. However, they would not be bothered to see feedback requests as long as they are not interrupting or too much distracting.

The consensus all external HCPs had was that feedback provision in search should be optional, and forced feedback provision may lead to invalid feedback. Moreover, as negative feedback accumulates, feedback collection requests might be annoying as negative feedback reflects the system performance to some extent.

HCP 1, "I guess neutral. The thing is that I'm not really looking to provide feedback on searches, right? I'm just trying to find the trial for a patient."

HCP 4, "Depends, in my case, it would depend on how much screen real estate you're using. If it's like half my screen, yes, I would be annoyed...If it's this size like one or two lines essentially, Then I wouldn't be too annoyed."

HCP 6, "It's sort of an option of giving feedback. You know, it's more like sometimes you're like, Okay, I'll give feedback. And sometimes you're like, Well, today, I don't have time to get feedback."

3) Feedback provision could be a way express to searchers' emotions

The attitudes of external HCPs have indicated that most of them feel neutral about feedback provision. In the cases where external HCPs are triggered to provide feedback, there is a certain emotional threshold to cross over (also found through interviews, p.68). For instance, to express gratitude by providing positive feedback when high relevance information is found, or to express

irritation by providing negative feedback for finding irrelevant information all the time.

HCP 5, "Ah, I normally don't use things like this. No, unless it's like, I cannot find something after looking at it for a long time then I'm a little bit irritated and like, no, it wasn't helpful, but normally should do stuff like this."

4) The value of feedback provision needs to be context fitting, sensible and practical

One important reason that hinders HCPs from always providing feedback is that they can not sense or do not know the values they would potentially gain from providing feedback. Some might be aware of the impact it could bring eventually, but those values cannot be presented to them immediately, or they could hardly streamline the work. It explains why more external HCPs preferred the first concept as the favorite because it brings immediate and practical values. In contrast, only one participant (HCP 9) explicitly regarded the virtual reward in concept three as of value, but all agreed that its embodiment does not fit the context.

HCP 4, "Indeed, as you mentioned somewhere that this is to optimize your search algorithm, you have something and that you have indeed a kind of reward scheduled and yeah, then people might do this."

HCP 6, "I think I wouldn't use it all the time. But if it really made a difference in my filtering, then I would use it."

5) Relevance feedback on the same information may differ

External HCPs were aware that the feedback they give might vary depending on their goals in Search. For example, the differences between doing research and finding eligible Clinical Trials for patients. Also, the relevance feedback on the same study might differ due to the differences (e.g. demographics and diagnosis) in patients.

HCP 7, "But I think if you were actually using the search engine to search for a fitting study for certain patients, and quite often you have to click that the study is not relevant because the patient does not fit the inclusion and exclusion criteria. So it's not the feedback that the study is not good. But it's the feedback that your patient does not fit the study."

6) Relevance is not just a yes and no question

Relevance Feedback is not merely a binary question (consistent to previous finding, p.66-67). In some cases, HCPs are unable to make a valid judgment because of insufficient information or time. Therefore, HCPs thought that the 'not sure' option should be kept, and the reasons behind should be collected. However, having 'not sure' in the first concept does not much make sense to some external HCPs because the feedback collected differs from the other two concepts that collect feedback after reviewing the content. Instead, it merely represents the likelihood of reading or not.

HCP 3, "I think we need it (the not sure option), since in some cases, we need such an option to re-examine them."

HCP 3, "Yes. Because then you can mark it and discuss it with your colleague or something like that. So it will be visible that the things I'm sure about are the things I'm not sure about to make the difference."

7) The usage of feedback should be communicated clearly

External HCPs were even more curious about how the feedback provided will be used than peers and internal HCPs. They wonder if the data will be for personal usage or system optimization, or even having an impact on the data source.

Moreover, HCP 4 and 5, explicitly expressed their worries that the relevance feedback provided in the system may lead to unwanted effects (e.g. filter bubble or 'polluting' the database). Some information regarded as irrelevant by the system may not be showing, which conflicts with HCPs' goal of finding a complete set of information.

HCP 2, "I am wondering, will my feedback just be used in your backend as data for the system usage, or just for me to record?"

HCP 4, "I might be a little bit worried that you can get quite a kind of the Facebook filter problem, like you click the things you like and in the end, you only see the things you like, and you'll get a tunnel vision."

HCP 5, "It's like just for yourself to go back later and see how I give the feedback or is it for PubMed, the feedback? I think it's very hard to give PubMed feedback because my research question can be maybe a lot different than someone else, and you can use the same term for that. If it's for myself, I think I would do it, maybe even more easily."

6.2 A guide for designing explicit feedback collection in text-based Search

The generalized guide was synthesized through the Design methods (Chapter 3) and insights gained from user studies (chapter 4, 5). Stemmed from relevance feedback collection for a Search in the healthcare context, the guide aims to help researchers, designers, or product teams develop interactions that engage end-users to provide user feedback in text-based Search.

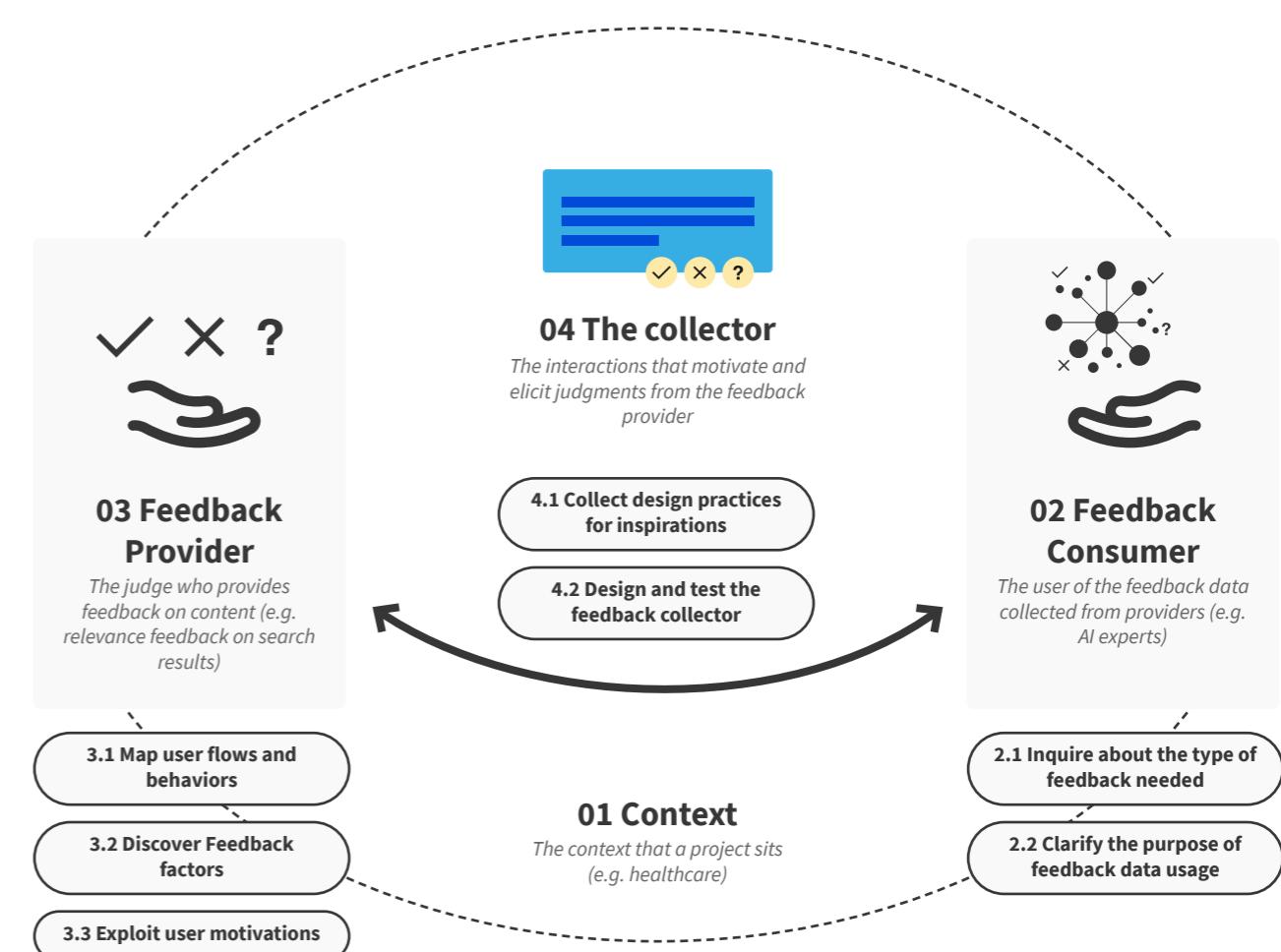


Figure 69. The relationship of four key factors and actions under each factors

6.2.1 A four-stage design guide

The guide adapts from the Human Centered Design process (IDEO) and proposes four-stage process and ten specific recommendations on the feedback collector design (the fourth stage).

Each stage features one key factor and Figure 69 shows their relationship and lists essential activities of each stage. The collector, with constraints in a specific context, plays an essential role of closing the loop between the feedback consumer and the provider.

1) Immerse the project context

The first stage aims to discover potential constraints through context immersion. Context, as the foundation, limits the space of design in nature (e.g. differences in data collection for adults and children). Better understanding the context (e.g. Contextmapping, Visser et. al, 2005) would contribute to better scoping the design space and inspirational resonations.

2) Understand the feedback consumer

The second stage aims to understand why feedback is needed, and how it will be used. The feedback consumer refers to the stakeholder that uses the feedback data for specific purposes (e.g. AI experts for evaluating or training the system, or product team for improving user experience). Researching the feedback consumer helps to identify the type of feedback needed and the purpose of its usage, which is necessary to be communicated to those who provide feedback. This stage includes two steps:

2.1 Inquire the type(s) of feedback needed: This activity is the starting point, and it is essential to elicit and understand what type of feedback is needed (e.g. relevance, satisfaction, qualitative, quantitative) as the design foundation.

2.2 Clear the purpose of feedback data usage: The other important point is to clear the purpose of feedback data usage (e.g. product analysis, feeding the algorithm) and how it will benefit the feedback provider within a particular time frame. It serves as information for deepening the understanding and is also essential to communicate to the provider for resolving potential concerns.

3) Determine and understand the feedback provider

The third stage aims to discover the constructing factors of feedback and find inspirational insights for designing the collector. The feedback provider means the judge (normally the end-user) who provides feedback. This stage features three key activities:

3.1 Map user flows and behaviors: In most cases, the feedback collector might be designed for an

existing digital product where user flows have been established and manifest behaviors patterns of certain user groups. Mapping the flows and understanding behaviors help to identify the point of engagement in providing feedback. It could be achieved by user observation or log analysis.

3.2 Discover factors that constitute feedback judgment:

What constitutes feedback judgment is essential, especially when predefined feedback options are required in the collector. To elicit the factors, researchers or designers could employ heuristics while users are interacting with the digital product (e.g. ask users to think aloud for the reasons of relevance a given in a Search).

3.3 Exploit user motivations: Motivations are the key for the provider to respond to the feedback collector. There are generally two types of motivations (intrinsic and extrinsic motivation), and it depends on the context, project objectives, and available resources to determine what motivators to be embedded in the feedback collector. This exploitation could be achieved through case studies of similar products and co-creation with focus groups.

4) Ideate and iterate the feedback collector

The last stage strives to design the embodiment of the feedback collector that engages and motivates the feedback provider. It comprises two key activities, and it is an iterative process requiring the involvement of the feedback providers to evaluate the design. More importantly, this stage synthesizes ten design recommendations categorized into three themes (Figure 73) as Fogg's behavior model for persuasive technology.

4.1 Collect design practices as inspirations:

In practice, there are a considerable amount of digital products collecting user feedback. By purpose, there are generally two types of feedback collectors: general feedback collector and specific feedback collector. The former aims for broad and fuzzy purposes. There are a lot of third-party services as a plugin (e.g. Usabilla, Hotjar). In comparison, the latter collects feedback data for particular purposes and is commonly found in recommendation systems (e.g. YouTube). Table 18 displays common design patterns in feedback collectors.

Purpose		
General	Specific	
Triggering position		
Fixed	Attached	Contextual
Collection methods		
Binary judgment	Scale	Predefined options
Free textbox	Pointing out	

Table 18. Categorizations of feedback collector design in practice

4.2 Design and test the feedback collector embodiment (Ten recommendations):

To consolidate the feedback collector design, an iterative process of designing and testing with target users is needed. For the design of the feedback collector, previous research has learned a list of ten recommendations (p.137-140) in three categories of Motivation, Ability, and Prompt:

Motivation In most cases, providing feedback is not the goal for end-users as it usually does not directly or immediately enhance what they want to achieve. Thus, employing proper motivators is essential.

Ability Motivators may not always be powerful enough to the feedback provider, and its efficacy differs among individuals. Hence, reducing the ability required from feedback providers could contribute to engagement.

Prompt The prompt manifests as the feedback collector UI, and the design of the collector impact how the feedback provider responds and the quality of feedback data.

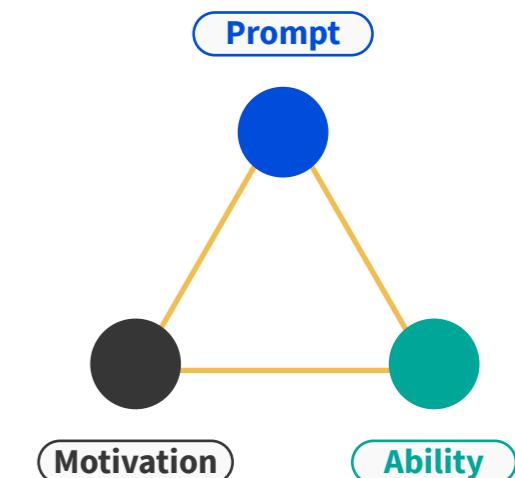
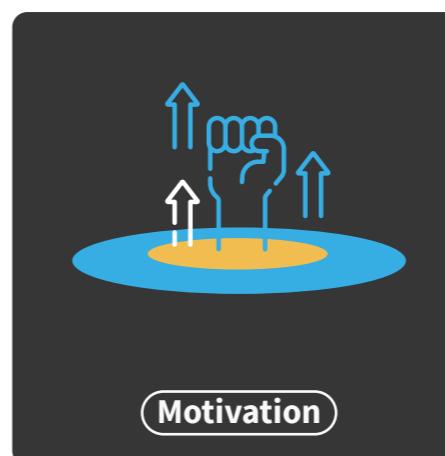


Figure 70. Three key themes in designing the feedback collector.



#1 Use motivators that fit the context

The motivators embedded in the feedback collector should match the context because inappropriate motivators may leave a false impression on end-users. For instance, using motivators that are too fun (enjoyment) may reduce the seriousness and professionalism of a search in a healthcare context.

Example: utility, enjoyment, reward, payoffs etc.

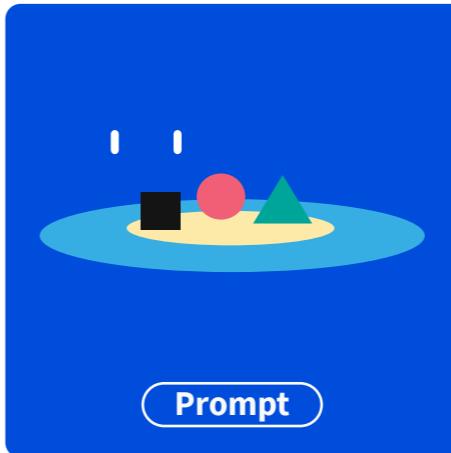


Motivation

#2 Communicate the usage of feedback data

Communication of how the feedback data will be used and how long it will come into effect is crucial as it increases transparency and resolves the potential concerns of end-users, such as filter bubbles. Also, it sets proper expectations for feedback providers.

Example: system optimization, personalization

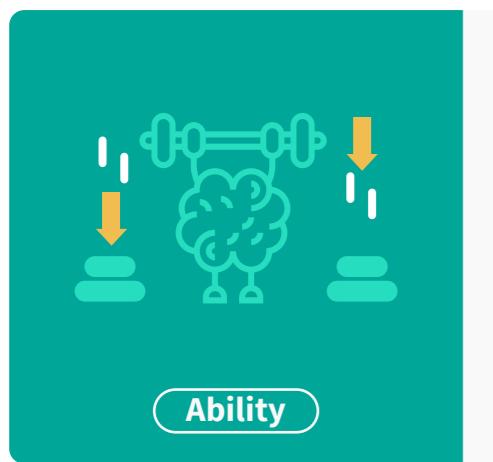


Prompt

#5 Align aesthetics to the context

Aesthetics creates the first impression, and it impacts how users perceive the whole system. The feedback collector should align aesthetics to the context.

Example: avoid using a cute embodiment in a serious context as healthcare



Ability

#3 Minimize perceived cognitive effort

To some extent, the perceived effort is equivalent to the perceived time required for achieving a task (providing feedback). The time during interactions is an effective indicator of cognitive involvement.

Example: display one feedback collection request at one time, show the time needed.

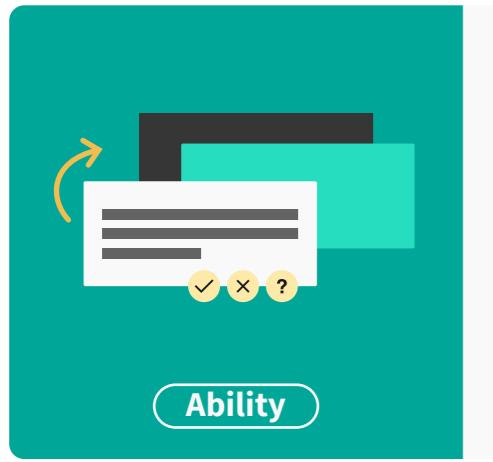


Prompt

#6 Nudge users at the right time

The 'right time' has a large room for explanation, but it is predictable and could be learned from research on the feedback provider. The higher confidence of timing the feedback collector prompts, the more likely the provider will be engaged, at least not to be bothered. Also, the quality of feedback may be better.

Example: ask for feedback at the moment after the provider has finished reading a search result



Ability

#4 Minimize interaction complexity

Minimizing the interaction complexity is a golden rule in User Experience in most cases (gaming experience might be exceptional), and the easier the interactions are, the more likely users would react to the feedback collectors.

Example: one click of a button



Prompt

#7 Balance the visibility and interruption

It is easy to design the collector that is apparent and attracts users' attention, but the difficulty is that it draws attention without creating interruption or distraction. For example, pop-ups attract much attention, but in most cases, it breaks the user flow. However, not being apparent would lead to ignorance.

Example: feedback request at the bottom of the content page could be easily missed, and avoid pop-ups

Key takeaways chapter 06



Prompt

#8 Keep the tone neutral

Keeping the tone neutral avoids stirring users' emotions, especially when providers give negative feedback. In an extreme case, when providers give negative feedback all the time, a neutral wording could avoid highlighting the dissatisfaction or unusefulness of a product.

Example: use the wording (for relevance feedback) 'relevant' or 'irrelevant' rather than 'useful' or 'useless'

6.1 Concept evaluation with target users

The concept evaluation involves nine healthcare professionals from three countries (Netherlands, China, and Brazil). The evaluation results indicate that the baseline search is perceived as a good product to use with an average SUS score of 75.31 (out of 100, Std=9.63).

On top of the baseline search, the first feedback concept (Pre-screener) received the highest average score (3.78/5, std=1.23), and more than half of the participants chose it as the favorite concept. However, the second concept (Reminder) appeared to be more stable (average score=3.56/5, std=0.83) and was regarded as the most motivating concept among all three by four participants. The last concept had the lowest experience score, and no one selected it as the favorite concept because all thought that it does not fit the context. However, most participants expressed their concerns that the first concept would not be able to collect well-considered feedback as it gathers feedback before users read the content. Thus, some suggested combining the first and second concepts, and one participant suggested bringing the enjoyment motivator to the first concept.



Prompt

#9 Give users the control to opt-out

Forcing users to provide feedback could potentially collect more feedback data, but it has the risk of annoying users or quitting. Moreover, the quality of the feedback data collected can be worse.

Example: always place a close or skip button, allow users to turn off feedback collection



Prompt

#10 Allow changes on given feedback

Feedback could change over time and is influenced by many factors. Thus, allowing changes in given feedback prevents mistakes and increases the quality.

Example: offer the action of changing feedback

6.2 A guide for designing explicit feedback collection

6.2 A guide for designing explicit feedback collection

The guide consists of a four-stage design cycle of "immerse the project context", "understand the feedback consumer", "determine and research on the feedback provider", and lastly "Ideate and iterate the feedback collector." All stages are interconnected, the context implies constraints which limit the design of the feedback collector that bridges the feedback provider and the

consumer. At each stage, except the first, there are specific steps to gain understanding and collect essential materials for designing the feedback collector.

The last stage features ten recommendations for designing the feedback collector embodiment, categorized into three themes of motivation, ability, and prompt following Fogg's behavior model of persuasive technology. The recommendations instruct the design and help avoid potential risks of disengagement:

- #1 Use motivators that fit the context;
- #2 Communicate the usage of feedback data;
- #3 Minimize perceived cognitive effort;
- #4 Minimize interaction complexity;
- #5 Align aesthetics to the context;
- #6 Nudge users at the right time;
- #7 Balance the visibility and interruption;
- #8 Keep the tone neutral;
- #9 Give users the control to opt-out;
- #10 Allow changes on given feedback.

07

Conclusion & Discussion

This final chapter includes the conclusions on answering the research question and a discussion of the project limitations, followed by the recommendations for myTomorrows and future work. Lastly, it closes with a personal reflection.

Chapter overview

- 7.1 Addressing the research question
- 7.2 Project contributions
- 7.3 Limitations and recommendations
- 7.4 Personal reflection

7.1 Addressing the research question

This section discusses the answers to the research question and implications of this research.

7.1.1 research question

How to engage healthcare professionals to provide trustworthy relevance feedback on search results in myTomorrows Search?

The purpose of this project was to use Human Centered Design methods to find solutions for engaging end-users (healthcare professionals) to provide relevance feedback on search results, and ultimately, using the feedback data for system optimization. In answering the central research question, this project went through the literature and exploratory research internally and externally. These insights laid the foundation for the three feedback collection concepts (chapter 5). The learnings from the process and evaluation sessions with peers, myTomorrows employees, and external HCPs formulated the guide for designing explicit feedback collection (chapter 6). There are three key factors to consider for HCPs engagement in providing explicit relevance feedback following Fogg's behavioral model. In a nutshell, to engage HCPs, it is needed to maximize their motivation, minimize the ability, and trigger HCPs with a proper prompt (feedback collector) design. The discussion follows the sequence of the sub-research questions.

RQ1: What are the incentives for HCPs to be engaged?

The theoretical research (Chapter 2) has mapped out identified multiple motivators (mainly in crowdsourcing) for engagement. In general, it could be divided into intrinsic and extrinsic motivations. To be more specific, it includes personal utility, altruism, enjoyment (symbolic rewards), etc. In this project, the three types mentioned above were incorporated in all three feedback concepts for comparison.

Throughout evaluations of all three concepts with external HCPs, personal utility (concept 1, a pre-screening tool helping users sort and manage search results) works better than the other two types because HCPs would be able to see the value immediately. However, HCPs perceive personal utility distinctively as the saying One man's trash is another man's treasure, vice versa. Personal enjoyment (an avatar with virtual badges as the reward) may work for some HCPs, but not all. In comparison, altruism is more universally accepted by most external HCPs.

RQ2: What type of interaction approach would better trigger HCPs to provide feedback?

The previous question has answered the incentives that motivate users to act, and this question probes into the prompt (i.e. the feedback collector). Throughout the studies on peers, internal and external HCPs, it is learned that an unobtrusive, easy, and direct feedback collection interaction that gently prompts at the right time could better trigger HCPs to provide relevance feedback. It should be as easy as possible, like pressing a button, and external HCPs don't expect too much fun or novelty in interactions as it may increase unnecessary complexity in providing feedback over time.

RQ3: At what moments should HCPs be nudged/motivated to provide feedback?

The studies of literature on the information-seeking process (Chapter 2), observations on how internal HCPs behave (Chapter 3), and external HCPs' heuristics on three feedback concepts (Chapter 6) identified two primary stages in which HCPs make relevance judgments: while HCPs scanning through the search result page (SERP), and when reviewing the content of results. In the speed dating stage, it was

also experimented to collect relevance feedback (Post hoc feedback, p.93) after HCPs have reviewed a few results, but it didn't take off due to the demanding cognitive efforts (requiring users to recall).

At SERP, it could collect relevance feedback on those irrelevant ones filtered out directly in the head. However, the main drawback is that such judgments could be superficial because they are made merely based on the study title and an introduction, which could mislead in some cases.

Within the stage of reviewing result content, the most suitable moment is the interval between HCPs have just finished reviewing the content of one document and the start of reviewing the next one(s). In this case, nudging users to provide relevance feedback would least interrupt user flows and requires the least perceived effort. However, the system should be able to distinguish

such a moment. Otherwise, it could cause interruptions to the reading experience.

RQ4: How to make sure the feedback provided by HCPs is trustworthy?

First, it needs to be addressed that false or misleading relevance feedback is inevitable due to individuals' mysterious human natures and situational cases. All participants in this research held authentic attitudes. Nevertheless, the findings through this research draw two key points to the answer.

First, HCPs need to be convinced that such data collected will be used for the right purposes (i.e. communication of the usage of feedback data). Second, the more passive HCPs are engaged in providing relevance feedback, the less trustworthy the feedback would be (i.e. not to force users to provide feedback).

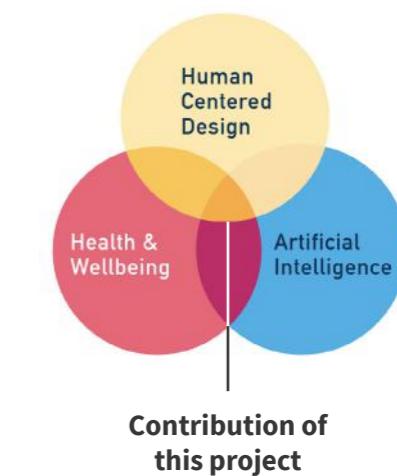
7.2 Project contributions

myTomorrows is the direct beneficiary of this research, besides, the contribution reaches the realms of the Design practice, and AI discipline.

7.2.1 Contribution to myTomorrows practice and Health & Wellbeing

As the research material, the conceptual search design enhanced with relevance feedback collection was evaluated by end-users. myTomorrows could use the tested design elements (or new features) for implementation in the actual treatment Search. Moreover, the insights gained from external HCPs could help the product team to iterate the current experience and reflect/envision future product strategies. Second, the generalized guide for designing explicit feedback collection. It could guide and inspire myTomorrows' designer(s) and product team to conduct certain activities, avoid potential risks, and mitigate biases when seeking solutions to engage users in providing relevance feedback.

myTomorrows, as a pharma-tech company, explores possibilities in the health & wellbeing domain. Another indirect contribution of the research project would be that the conceptual Search offers a new design perspective on how investigational treatment information could be more accessible to end-users and inform a better clinical decision making.



7.2.2 Contribution to Design practice

To the best knowledge, designing interactions for explicit relevance feedback collection in practice was rare and non-systematic. Publications around relevance feedback are focusing on collecting and exploitation of implicit feedback. Nevertheless, there is one guide (Google people +AI guidebook) for designing feedback collection within the AI discipline in practice. However, it merely consists of key considerations without a step by step guide of what actions to take to achieve the goal, and the scope is both on implicit and explicit feedback collection.

In comparison, this guide generalized through the research focuses on explicit user feedback collection. It provides a step by step guide adapted from Human Centered Design methods and lists ten theoretically and practically grounded (only in the context of Search for investigational treatment options at this stage) recommendations to

which user researchers or designers could refer. Stemmed from the research in the particular context of AI (IR and recommender) and health & well-being, this guide could potentially apply to similar contexts (e.g. academic search systems) where relevance is a key metric and could impact a broader audience possibly.

7.2.3 Contribution to AI discipline

Compared to using implicit feedback that is inherently available, collecting explicit feedback might not be the ideal approach (chapter 2, p.40). However, in some cases like myTomorrows, the limited user scale forces the AI team to turn to explicit relevance feedback, even though it is aware of the risk of interrupting end-users. Therefore, the outcome (the generalized guide) could offer an alternative opportunity to AI teams who want to use feedback data from users for improving systems in similar situations.

a solid theoretical foundation for the exploratory research. However, these fields of studies have a considerable amount of literature, and within the given timeframe, some information or insights were inevitably missed.

Limitations in the exploratory research

The target user of the project is external HCPs. Ideally, all activities involved human participants were supposed to be conducted with the target group. However, due to the limited resources and unavailability of external HCPs in most cases, it was not possible to involve them in all research activities (i.e. Heuristic evaluation and search behavior observation). Therefore, the results might not be representative of the actual target groups. Additionally, collecting examples of explicit feedback collection in practice may not be thorough, limiting the inspirational scope.

Limitations in the concept design and prototype

The concept design was a synthesis of the insights from literature, exploratory research, and brainstorming ideas with peers. However, external HCPs were not directly involved in the brainstorming sessions, leading to the lack of early input to the concept from the target group. Moreover, another limitation was that the prototype of the concept could not function to retrieve information due to technical boundaries.

Limitations in the test and evaluation

Luckily, the test and evaluation sessions successfully recruited nine target end-users. However, the limitations in the design prototype used for evaluation restricted external HCPs' behaviors in predefined flows where 'COVID-19' was the only query used for retrieving treatment options. It could have brought an inferior experience to those whose expertise is not on virology. Hence, inferior qualitative data were collected.

Besides, all studies with external HCPs were no longer than one hour per person, and it lacked longitudinal study results in an actual working setting.

Limitations in the generalized design guide

The guide for designing explicit feedback collection was generalized from the insights gained through user studies with the nine feedback ideas (p.103-104) and three feedback concepts (p.134-136). Also, it blended in the learnings from the literature and the project process. However, the guide was not validated with the target audience (user researchers or designers). Thus, it remained uncertain how useful it could be and in what particular fields this design guide could be applied to.

7.3.2 Recommendations for development at myTomorrows

The three feedback concepts (p.116-121) have laid the foundation for implementation. However, there remains a distance from production. For myTomorrows to implement explicit relevance feedback collection interactions in the treatment Search, the recommendation is to combine each concept's essentials and follow the generalized guide's ten recommendations (p.139-142). The first concept (Pre-screener) should be the foundation, aided with the second concept (Reminder) when no feedback could be collected from the search engine result page (SERP). Also, the virtual reward (e.g. badge) from the third concept (Hearty) could be incorporated but unobtrusively and in a more serious manner. In essence, explicit feedback collection in myTomorrows Search should be easy to interact with and try to become a utility tool. Additionally, the fused recommendation mentioned above could be brought only to run A/B test to gather quantitative data.

7.3 Limitations and recommendations

This section discusses the limitations of the project by stage and proposes recommendations for myTomorrows of what to implement and future research around designing feedback collection.

7.3.1 Limitations of this research

This research was conducted during the COVID-19 pandemic, meaning that nearly all activities were conducted online. It could have an unmeasurable negative impact on the whole project, especially on the studies that involve human participants (i.e. brainstorming sessions and evaluation sessions with HCPs). Except for the force majeure, there are other limitations in the project. By stage, those limitations stem primarily from literature research

(chapter 2), exploratory research (chapter 3), design and prototype (chapter 4), test and evaluation (chapter 5), as well as the generalized design guide (chapter 5).

Limitations in the literature research

The research into literature was done in the fields of Information Retrieval, search behaviors, relevance feedback, user engagement, and motivational models (e.g. nudging and persuasive technology). The insights from this research laid

7.3.3 Recommendations for future research

This research aimed to find solutions to engage healthcare professionals to provide explicit relevance feedback on search results. Such research is scarce and based on the generalized

guide, a few future research directions might be worth exploring as below:

1) Research on the usefulness and effectiveness of the guide

Apply the design guide to different contexts to evaluate its usefulness and effectiveness of the design process and the ten recommendations.

2) Research and create a motivational design framework

Providing user feedback requires the right motivations. However, this research didn't focus much on mapping out all motivational factors.

3) Research on communicating feedback usage

This research merely tried to communicate the usage of feedback through words. Still, there might be other approaches (e.g. gamification) to achieve so in a more engaging way and with educational purposes.

7.4 personal reflection

This project began with a short but rather concrete brief titled Clear and Effective Interactive Search Experience Design. Having little knowledge about Search, I started to immerse myself in the knowledge world of Search (IR). Books like Introduction to Information Retrieval, Search User Interface Design, and Search Patterns helped me form a holistic overview of what Search is from technical and design perspectives. The knowledge gained from the readings revealed an iceberg to me and was overwhelming. For a Search product, users see so little, but there are a considerable amount of interaction details to consider. I have to say that search experience is a great successor of the philosophy 'Less is More', and this was my primary motivation to choose this topic as my Master Thesis.

However, understanding Search foundations is merely the starting point of the project. What comes essential is the question 'how to motivate end-users to provide relevance feedback on search results so that it could be used for system-level optimization'. Initially, I had the assumption that users would be reluctant to providing feedback, which was proven through literature and studies with end-users. Therefore, there was a lot of research around engaging experience and motivational theories (e.g. Nudging and Persuasive technology). These theories, to me, is another iceberg in the knowledge sea that requires more in-depth investigations.

Nevertheless, there is another iceberg I encountered during the project: healthcare and medical discoveries. As I talked to experts in the health domain and put myself in the daily exposure to the news and information around medical advances (mainly following COVID-19), I gradually grew a deeper understanding of how the current medical discovery process needs a dramatic change and how this project could potentially save the lives of many.

After six months of research at the intersection

of AI (IR and recommender), health & well-being, and Human Centered Design, I have gained a robust understanding of each aspect of this project, especially designing feedback collectors. A long time exposure to this project has made me excessively sensitive to the term Feedback and critical about feedback collectors in practice, which reflects my effort put in the project.

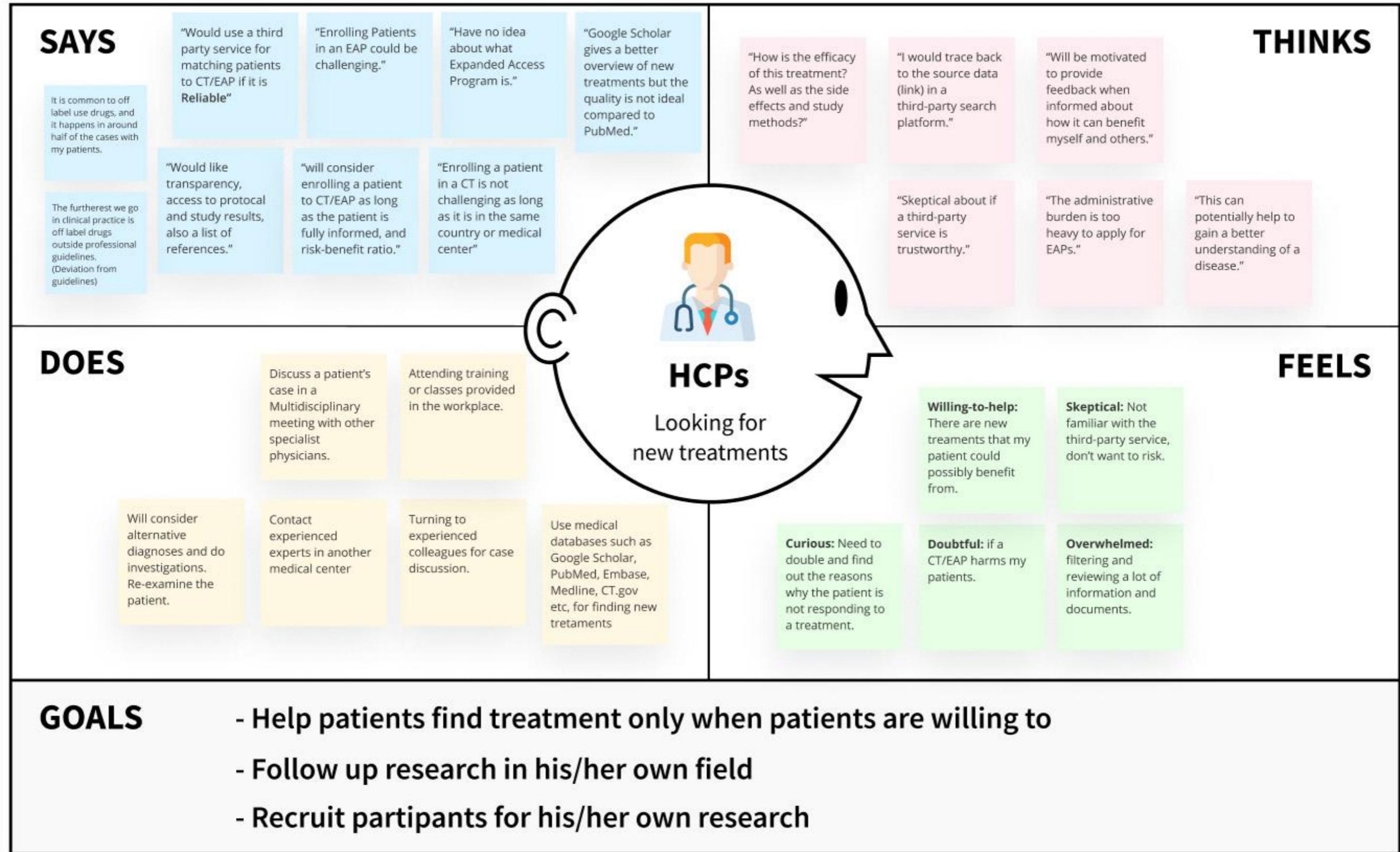
During the project, I worked closely with the internal designer Melissa and internal HCPs who help patients with unmet medical needs generate treatment search reports. My concept (baseline) design successfully stirred the interest of the medical and the business department. Moreover, some ideas from the baseline concept were already under consideration for implementation.

The main personal learning from this project is that I am more aware of scoping a project. From the beginning to the end of the project, the scope and approach changed multiple times. It was a frustrating experience, but it is essentially the resilience We designers must equip. The other I learned is about clarifying ideas. In most cases, We designers (or maybe just I) tend to use fuzzy words, which may sound good but leaves a large room for the audience to interpret, potentially leading to misunderstanding. Overall, I enjoyed working on this project with the supervisory team and the colleagues at myTomorrows. I firmly believe that I will take away this unique learning experience with me for long.

Bibliography

- Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011, February). Towards a science of user engagement (position paper). In WSDM workshop on user modelling for Web applications (pp. 9-12).
- Bajic, D., & Lyons, K. (2011). Leveraging social media to gather user feedback for software development. Proceeding of the 2nd International Workshop on Web 2.0 for Software Engineering - Web2SE '11, 1-6. <https://doi.org/10.1145/1984701.1984702>
- Bates, M. J. (1989). The Design of Browsing and Berrypicking Techniques. <https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html>
- Broder, A. (2002). A taxonomy of web search. ACM SIGIR Forum, 36(2), 3. <https://doi.org/10.1145/792550.792552>
- Budiu, R. (2019, November 10). Information Foraging: A Theory of How People Navigate on the Web. <https://www.nngroup.com/articles/information-foraging/>
- Bunnik, E. M., & Aarts, N. (2019a). What do patients with unmet medical needs want? A qualitative study of patients' views and experiences with expanded access to unapproved, investigational treatments in the Netherlands. *BMC Medical Ethics*, 20(1). <https://doi.org/10.1186/s12910-019-0420-8>
- Bunnik, E. M., & Aarts, N. (2019b). What do patients with unmet medical needs want? A qualitative study of patients' views and experiences with expanded access to unapproved, investigational treatments in the Netherlands. *BMC Medical Ethics*, 20(1). <https://doi.org/10.1186/s12910-019-0420-8>
- Bunnik, E. M., Aarts, N., & van de Vathorst, S. (2018). Little to lose and no other options: Ethical issues in efforts to facilitate expanded access to investigational drugs. *Health Policy*, 122(9), 977-983. <https://doi.org/10.1016/j.healthpol.2018.06.005>
- Caplan, A. L., & Bateman-House, A. (2015). Should patients in need be given access to experimental drugs? *Expert Opinion on Pharmacotherapy*, 16(9), 1275-1279. <https://doi.org/10.1517/14656566.2015.1046837>
- Chopra, K., & Wallace, W. A. (2003). Trust in electronic environments. 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The. <https://doi.org/10.1109/hicss.2003.1174902>
- Darrow, J. J., Sarpatwari, A., Avorn, J., & Kesselheim, A. S. (2015). Practical, Legal, and Ethical Issues in Expanded Access to Investigational Drugs. *New England Journal of Medicine*, 372(3), 279-286. <https://doi.org/10.1056/nejmhe1409465>
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105-115. <https://doi.org/10.1037/h0030644>
- DiMasi, J. A. (2002). The Value of Improving the Productivity of the Drug Development Process. *PharmacoEconomics*, 20(Supplement 3), 1-10. <https://doi.org/10.2165/00019053-200220003-00001>
- DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, 22(2), 151-185. [https://doi.org/10.1016/s0167-6296\(02\)00126-1](https://doi.org/10.1016/s0167-6296(02)00126-1)
- Fogg, B. J., Swani, P., Treinen, M., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., & Shon, J. (2001). What makes Web sites credible? Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '01. <https://doi.org/10.1145/365024.365037>
- Furnas, G. W. (1997). Effective view navigation. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '97. <https://doi.org/10.1145/258549.258800>
- Gil, Y., & Artz, D. (2006). Towards content trust of web resources. Proceedings of the 15th International Conference on World Wide Web - WWW '06. <https://doi.org/10.1145/1135777.1135861>
- Guha, R., McCool, R., & Miller, E. (2003). Semantic search. Proceedings of the Twelfth International Conference on World Wide Web - WWW '03. <https://doi.org/10.1145/775152.775250>
- Haidich, A.-B., & Ioannidis, J. P. A. (2001). Patterns of patient enrollment in randomized controlled trials. *Journal of Clinical Epidemiology*, 54(9), 877-883. [https://doi.org/10.1016/s0895-4356\(01\)00353-5](https://doi.org/10.1016/s0895-4356(01)00353-5)
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. 2014 47th Hawaii International Conference on System Sciences, 3025-3034. <https://doi.org/10.1109/hicss.2014.377>
- Irvine, J. (2016, April 23). How a Physician Can Work With a Not Yet Approved Drug Through Compassionate Use. <https://thehealthcareblog.com/blog/2016/04/17/far-from-evidence-based-prescribing-the-world-of-compassionate-use/>
- Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, 6, 89-106. <https://doi.org/10.1016/j.invent.2016.10.002>
- Korth, H. F., & Silberschatz, A. (1997). Database research faces the information explosion. *Communications of the ACM*, 40(2), 139-142. <https://doi.org/10.1145/253671.253760>
- Marchionini, G. (1989). Information seeking strategies of novices using a full text electronic encyclopedia. *Journal of the American Society for Information Science*. [https://doi.org/10.1002/\(SICI\)1097-4571\(198901](https://doi.org/10.1002/(SICI)1097-4571(198901)
- Marchionini, G., & White, R. (2007). Find What You Need, Understand What You Find. *International Journal of Human-Computer Interaction*, 23(3), 205-237. <https://doi.org/10.1080/10447310701702352>
- Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds." *Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '09*. <https://doi.org/10.1145/1600150.1600175>
- Mehta, A. (2008). Gaucher disease: unmet treatment needs. *Acta Paediatrica*, 97(s457), 83-87. <https://doi.org/10.1111/j.1651-2227.2008.00653.x>
- Morville, P. (2010). Search patterns: design for discovery. O'Reilly Media, Inc.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938-955. <https://doi.org/10.1002/asi.20801>
- Penberthy, L. T., Dahman, B. A., Petkov, V. I., & DeShazo, J. P. (2012). Effort Required in Eligibility Screening for Clinical Trials. *Journal of Oncology Practice*, 8(6), 365-370. <https://doi.org/10.1200/jop.2012.000646>
- People + AI Guidebook. (n.d.). Retrieved April 7, 2020, from <https://pair.withgoogle.com/chapter/feedback-controls/>
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643-675. <https://doi.org/10.1037/0033-295x.106.4.643>
- Pressler, T. R., Yen, P.-Y., Ding, J., Liu, J., Embi, P. J., & Payne, P. R. O. (2012). Computational challenges and human factors influencing the design and use of clinical research participant eligibility pre-screening tools. *BMC Medical Informatics and Decision Making*, 12(1). <https://doi.org/10.1186/1472-6947-12-47>
- Schatz, B. R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net. *Science*, 275(5298), 327-334. <https://doi.org/10.1126/science.275.5298.327>
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/s0140525x00005756>
- Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search. *D-Lib Magazine*, 3(1). <https://doi.org/10.1045/january97-shneiderman>
- Sutcliffe, A., & Ennis, M. (1998a). Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10(3), 321-351. [https://doi.org/10.1016/s0953-5438\(98\)00013-7](https://doi.org/10.1016/s0953-5438(98)00013-7)
- Sutcliffe, A., & Ennis, M. (1998b). Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10(3), 321-351. [https://doi.org/10.1016/s0953-5438\(98\)00013-7](https://doi.org/10.1016/s0953-5438(98)00013-7)

Appendices



Search criteria

Condition: Type and select disease
Leave empty to search worldwide.

Sex of patient: Male Female All

Search home

1. Search
Enter the condition for which you want to find treatment options.
Note: Leave empty to search worldwide.

2. Review
Examine search results to determine patient eligibility based on criteria like sex, age and location. There are also filters you can use to narrow your search results.

3. Discuss
Discuss search results with a member of the myTomorrows medical team.



myTomorrows

Search as: Patient Healthcare professional

Search criteria: Condition: Cancer
Leave empty to search worldwide.
Sex of patient: Male Female All

Search filters: CT Clinical Trials: Phase 1, Phase 2, Phase 3, Phase 4, Not applicable; EAP Expanded Access: Available, Not yet recruiting, Completed.

One moment please... Searching database for your treatment options

Loading

myTomorrows

Search as: Patient Healthcare professional

Search criteria: Condition: Coronavirus
Leave empty to search worldwide.
Sex of patient: Male Female All

Search filters: CT Clinical Trials: Phase 1, Phase 2, Phase 3, Phase 4, Not applicable; EAP Expanded Access: Recruiting, Available, Not yet recruiting, Completed.

Process graph

43 results based on your search criteria and filters. View saved results(0)

Search within these results: Enter a term or intervention.

Type	Phase	Status	Title
CT	Recruiting	The Efficacy of Lopinavir Plus Ritonavir and Arbidol Against Novel Coronavirus Infection	
Interventions:	Lopinavir Tablets Ritonavir arbidol	Conditions: Communicable Diseases Coronavirus Infection Infection Severe Acute Respiratory Syndrome	
The study explores the efficacy of lopinavir plus ritonavir and arbidol in treating with novel coronavirus infection. As a result this study would provide evidence for the clinical usage of these drugs in the future.			
Print result More info			
CT	Recruiting	Evaluation of Ganciclovir Combined With Ritonavir in the Treatment of Novel Coronavirus Infection	
CT	Not yet recruiting	The Clinical Study of Camimycin on Treatment Patients With COVID-19	
CT	Recruiting	A Study to Evaluate the Efficacy and Safety of Pivendipine With Novel Coronavirus Infection	

Result list

myTomorrows

Search as: Patient Healthcare professional

Search criteria: Condition: Coronavirus
Leave empty to search worldwide.
Sex of patient: Male Female All

Search filters: CT Clinical Trials: Phase 1, Phase 2, Phase 3, Phase 4.

Result content

Clinical Trial | Phase: IV | Status: Recruiting | Start date: 28-01-2020 | End date (est.): 31-07-2020

The Efficacy of Lopinavir Plus Ritonavir and Arbidol Against Novel Coronavirus Infection

Purpose:
The study explores the efficacy of lopinavir plus ritonavir and arbidol in treating with novel coronavirus infection. As a result this study would provide evidence for the clinical usage of these drugs in the future.

Conditions: Communicable Diseases Coronavirus Infection Infection Severe Acute Respiratory Syndrome

Interventions: Lopinavir Tablets Ritonavir arbidol

Countries: China See more

myTomorrow search UX issues



Search home (3)

- Visuals in Patient/HCP are not consistent
- Information provided is irrelevant to search
- Information distracts users

Patient/HCP (1 + 2)

- Switching search mode restarts search (3/12)
- Explanations are not consistent compared to other ?
- Comparison and explanation of Patient/HCP are not clear (2/12)

Process graph (4)

- It does not correspond correctly to what users are doing (4/12)
- It does not indicate how to act next (3/12)
- The graph is different in search result page than in home page
- It distracts and confuses users (3/12)

Search box (8 + 4)

- Customized search query is not supported (2/12)
- Occasional autofill after hitting ENTER (2/12)
- It suggest conditions that have no results
- Type and search is not supported (8/12)
- It takes long to show suggestions (eg.Pancreatic cancer)
- It doesn't inform user when suggestion is not selected, and returns no result.
- "Type and select..." works the same as "Type..."
- The suggestions don't indicate scrolling
- Left aligned search button is not visually pleasant
- It doesn't support advanced search
- has no one-click clear (content)
- has no support for multiple-country search at once

Key insights from first impression

Looks clean and information is well organized (8/12, #1, 2, 3, 4, 5, 6, 7, 11)

The threshold for Patient to search is too high (5/12, #2, 3, 5, 9, 12)

It doesn't feel like a ready product, but an internal tool or prototype (4/12, #1, 2, 3, 11)

Some interactions are out of expectations, and the experience is overwhelming (6/12, #3, 4, 5, 7, 9, 11)

Disconnected from myTomorrows service (1/12, #1)

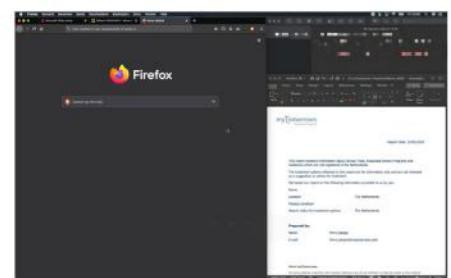
Visual styles are not consistent in different pages (2/12, #8, 9)

* Design (9) #1, 2, 3, 5, 6, 7, 8, 9, 10 / Linguistics (1) #4 / Medical science (1) #11 / Civil engineering (1) #12

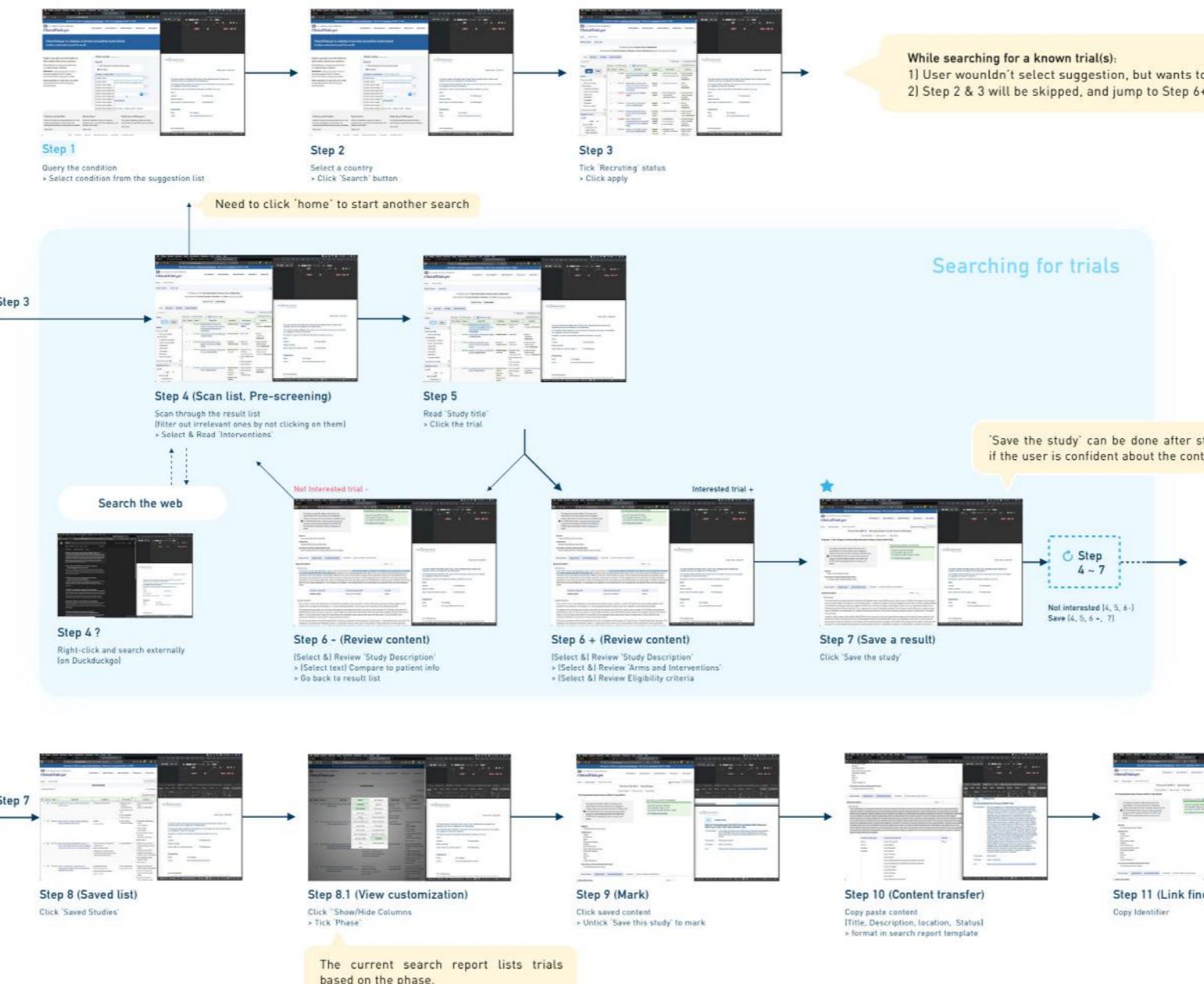
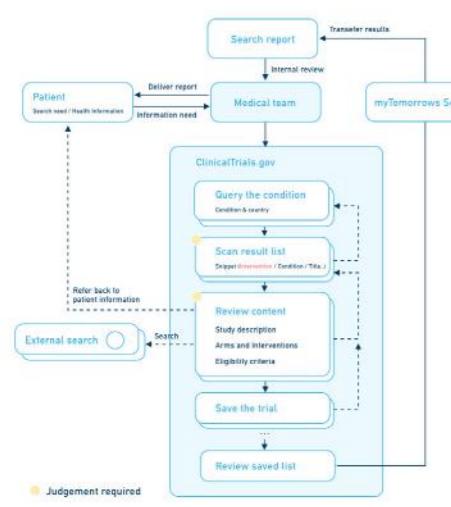
Current Search flow

From a search need to report generation

04 MAR 2020



Desk setup



HCP SURVEY:

⇒ 10 Survey Respondents

- 6 external HCPs
- 4 internal HCPs

⇒ Medical expertise overall

2. How many years of experience do you have working in a medical profession?

[More Details](#)

- | | |
|--------------------|---|
| Less than 5 years | 3 |
| 5 to 10 years | 3 |
| More than 10 years | 4 |



Breakdown between internal & external:

External HCPs (6)

- 3 x < 5yrs medical experience
- 3 x > 10yrs medical experience

Internal HCPs (4)

- 3 x 5-10yrs medical experience
- 1 x >10yrs medical experience

⇒ Recruitment efforts

- 3 x external HCPs would like feedback about the outcomes of this research.
- 5 x external HCPs consented to be contacted for future research initiatives.

⇒ General feedback

"The scenarios are not very specific. This can be improved, otherwise the answers that you will get will probably be very variable and may not be the answers you are looking for."

"Would have liked to be provided with some context.. What this questionnaire is about.. The purpose of it."

Insights from these final comments:

Question 3

Scenario: Sam has been diagnosed with a chronic condition and you are the treating specialist physician. You have followed the treatment guidelines for the condition, but Sam has not responded well to the last line of treatment available.

What would you do next in the management of Sam?

EXTERNAL

Will consider alternative diagnoses and do necessary investigations.
Re-examine the patient in case missed something. Check on history.
If find nothing consult a colleague or seek second opinion.

Find a clinical trial for inclusion

Discuss Sam's case in a Multi disciplining meeting with other specialist physician.

See if there is a combination possible and of course find out if there is a specific reason for not responding. If no options left, see if there is any new trial medication.

In the case that I would not be the best-known expert on The Netherlands in his condition, I would first contact an expert in another center for advice.

Discuss a second opinion

INTERNAL

Discuss this case on a higher level. IE: with colleagues in our journal club/with the prof of the department etc

Consult a trusted senior colleague. Quite likely in person or over the phone.

1. Check with the patient what they want and need.
2. Search on pubmed for latest treatment options and Clinical trial.gov
3. Discuss case with colleagues

Check clinical trials in own hospital

Question 4

Scenario: Sam joins a local support group and finds out there may be a treatment in development that could improve the quality of life for people with the same condition. Sam approaches you to find out if this is a possibility.

Would you consider enrolling a patient, like Sam, in a CT or EAP as a possible treatment option?

EXTERNAL

Important is to find out how reliable Sam's source is. If it is reliable, then a discussion with Sam should start to determine the pro's and con's of such a CT.
If not reliable, I will tell that to Sam as well.

Yes, as long as he is fully informed and no major risk to his condition
Yes, because nothing else is available and in both circumstances monitoring and sharing data is possible
Yes, I would consider it.
Yes, depending on the results and the possibilities, I would consider it.
yes, after I did my research. This also depends on the severity of the condition.
INTERNAL
yes, if the option is appropriate
Yes, if the data is promising, the safety profile is good (risk-benefit ratio) and if the patient is keen.
This would be very dependent on the disease and Sam's condition... don't feel there is enough here for me to give a properly reasoned answer.
I would because I have my patients best interest at heart. I would be apprehensive because of the lack of time and administrative burden

Question 5

Scenario: You advise Sam that you will research which treatments in development are available and plan a consult to discuss the options you find next week.

Do you think it would be challenging to enrol Sam in a CT or an EAP if this was a treatment option?

EXTERNAL
Sure.
I have never done this before. I assume it is very difficult since these trials usually have very strict rules?
See above (Yes, because nothing else is available and in both circumstances monitoring and sharing data is possible)
I am not clear what exactly is meant by expanded access programme
I think a clinical trial should be feasible, I don't know about EAP
Whether enrolling someone in a CT is possible is dependent on many factors, first of all is someone meets the inclusion criteria. I would not say this is 'challenging' as the answer to this question is yes or no. If the CT is not available in my center, I will refer the patient to another center. Things become challenging if there are only centers abroad that conduct the trial. I think depending on your personal contacts and experience, enrolling someone in an EAP can be challenging, but I have little to no experience with this.

INTERNAL
yes, due to inclusion criteria misfit in CT, or regulatory approval difficulties in EAP
depends on the situation, in own hospital CT is not difficult. EA takes some paperwork.
Yes, time and knowledge would be barriers to enrolling Sam in a CT/EAP.
yes. i have limited time. many times the CT PIs are non responsive. i dont have time to follow up on these emails to get sam enrolled.

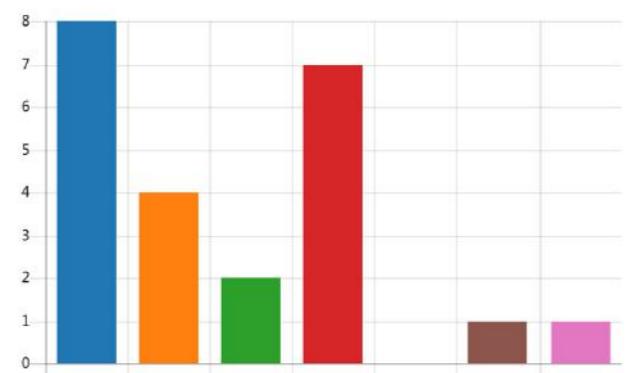
Question 6

Scenario: You get to work early and have time before your clinic starts and decide to start researching treatment options for Sam.

Which online platform would you use to search for treatment options for Sam?

Responses:

- Pubmed 8
- UpToDate 4
- Medscape 2
- ClinicalTrials.gov (CT.gov) 7
- CT Search Platform (Datasource...) 0
- CT Search Platform (Datasource...) 1
- Other (e.g. specific CT registry,...) 1



Additional comment

Q6- Other: Webzines may also be consulted first.

I was wondering where other members of the MDT fit into our research (the HCP research process). A pharmacist or nurse as the one referring the physician to (CT search) platform for example?

Findings & Insights

Question 7

Scenario: You find a few options, but you need to stop researching because it's time to start your clinic. You will have to continue when you have more free time.

Would you use a service where specialised doctors match a specific patient profile, like Sam, to Clinical Trials based on their eligibility?

EXTERNAL
Please specify 'a service'. I would consider using 'a service' when I know which one to use, when I know what this service does and when I have heard good stories about using 'this service'.
No not really.. Not familiar with this approach
Yes, when there are reliable services
Do not know where to find those
If that independent service exists.
If this is possible, I might consider this
INTERNAL
Yes
Yes, only if it is for free, does not take too long and not linked to any big pharma company/ biased.
yes. time saving. no need to search myself
If it is a trusted source without any obvious conflicts of interest, yes I would.

Findings & Insights

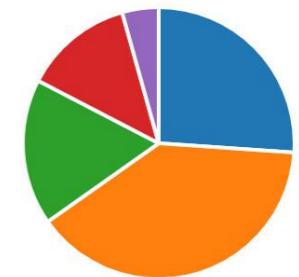
Question 8

Scenario: When asking a colleague for advice about Sam's case, he tells you about a CT Search Platform that incorporates multiple registries into its database. You decide to try it out.

What would help you trust the CT Search Platform as a credible source of information?

Responses:

- Indicates the registries include... 6
- Provides links to the data sour... 9
- Provides information about th... 4
- An FAQ explaining how to enr... 3
- None of the above 1



Additional comment

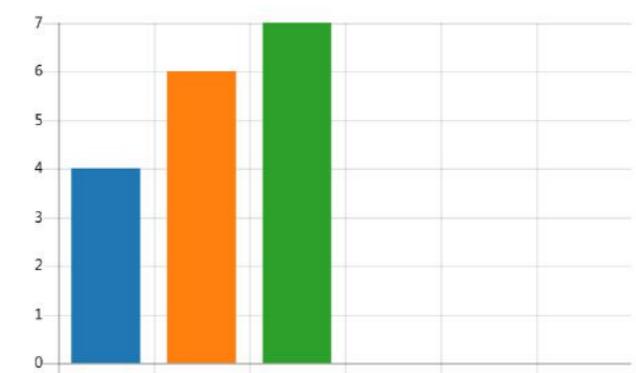
Q 8 I found hard to answer, since I have no experience with this. As a young MD I would like transparency, access to the protocol and if possible preliminary results, also a list of references on which the research is based on would be nice.

Question 10

Artificial Intelligence (AI) empowered search platforms require feedback from the people using them to continuously improve the platform's functionality. As a healthcare professional, what would encourage you to provide feedback on the search results you would obtain on such a platform?

Responses:

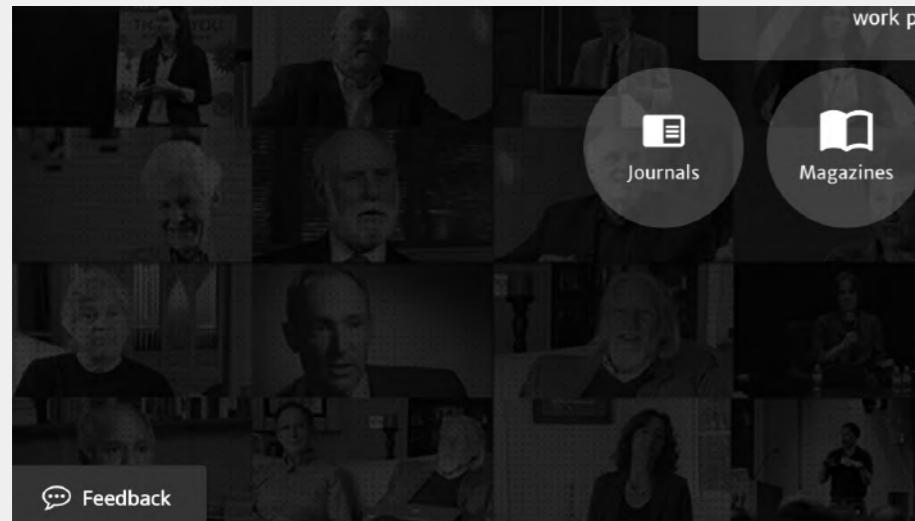
- Having the ability to remove a... 4
- Having the ability to save a re... 6
- Being informed about how pr... 7
- Receiving a virtual reward for ... 0
- Participating in a competition ... 0
- I would not provide any feedb... 0



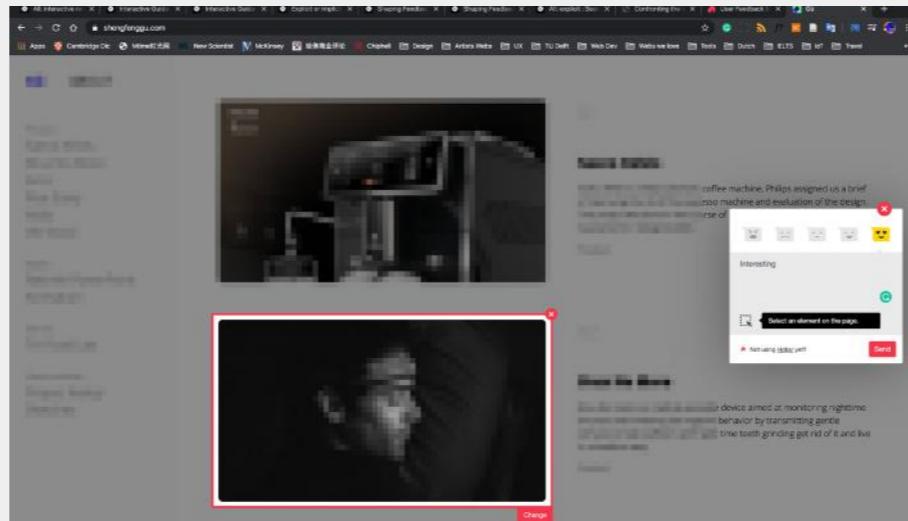
Findings & Insights

Feedback collection in practice

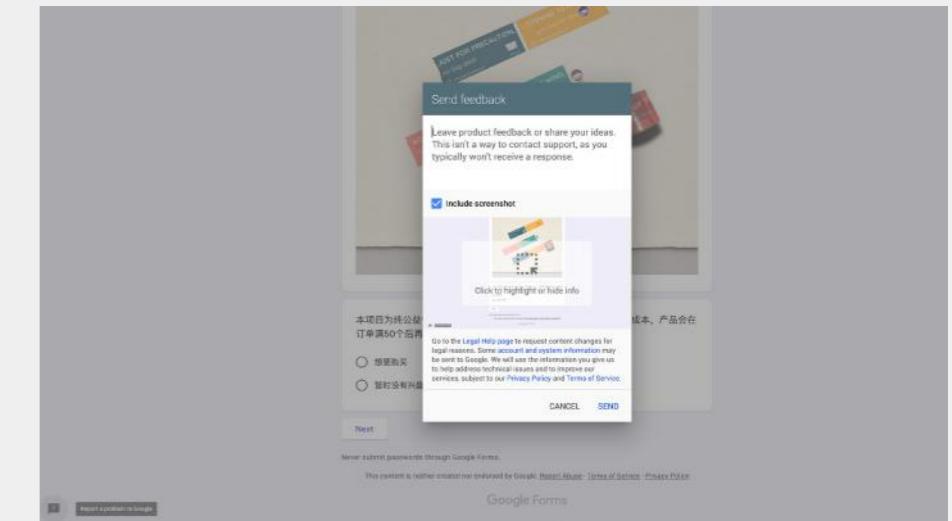
ACM digital library



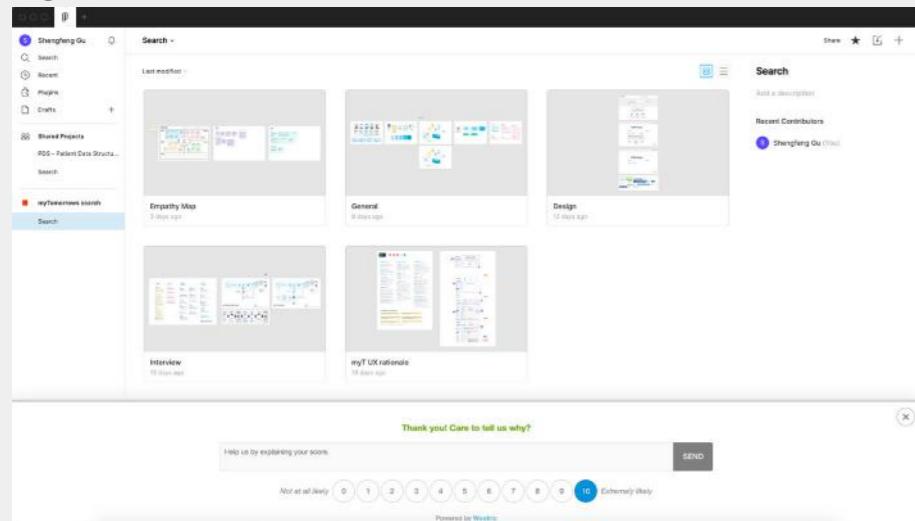
Hotjar



Google forms

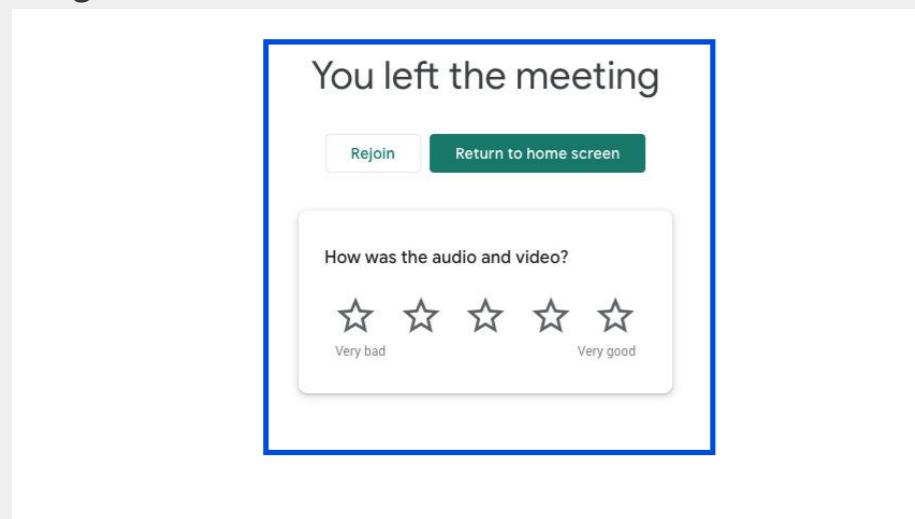


Figma

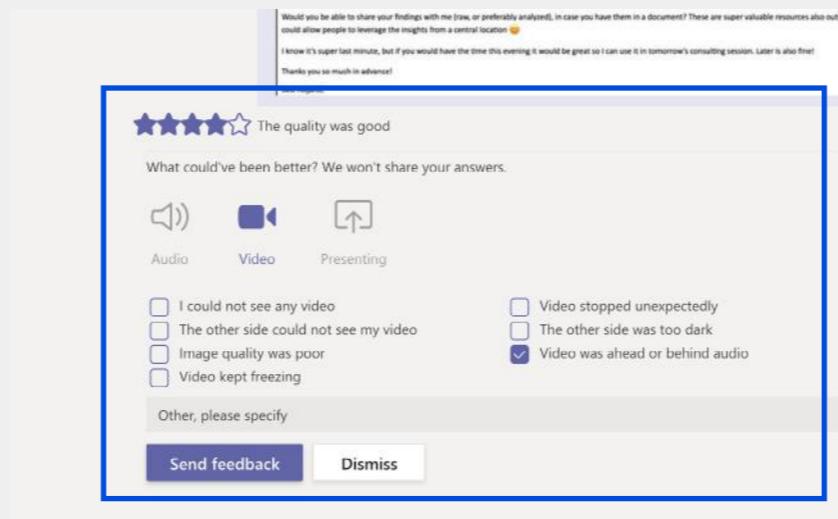


General feedback collection

Google meet

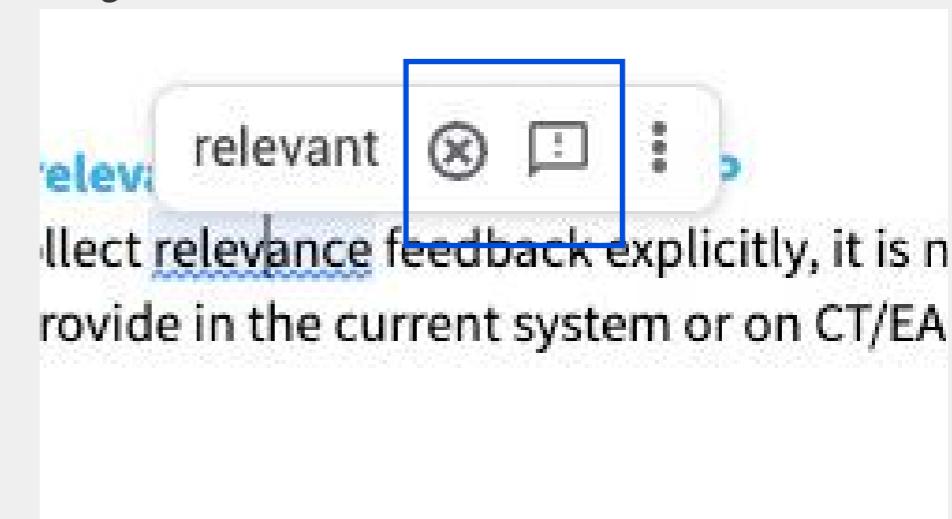


Teams



Specific feedback collection - Contextual

Google doc



Amazon

A screenshot of the Amazon product search interface. At the top, there are navigation links for 'Account & Lists', '& Orders', 'Prime', and 'Basket'. Below this, categories like 'Computers', 'Sports & Outdoors', and 'Pantry' are listed. A message 'Covid-19: Delivery, returns & safety' is displayed. In the center, a search bar has 'Sort by: Featured' selected. At the bottom right, there is a blue-bordered 'Ad feedback' button.

Google photo

A screenshot of the Google Photo interface. It shows two circular profile pictures of a young child. Below each picture is a green checkmark icon, a red circle with a slash icon, and a question mark icon. The words 'Same', 'Different', and 'Not sure' are centered under their respective icons. The background features a dark, slightly blurred image of a person.

Google help

A screenshot of the Google Help interface. It includes sections for managing privacy settings for ads, turning off ads, and filling in lead forms. A specific feedback form is highlighted with a blue border. It asks for feedback about an article and includes 'Was this helpful?' buttons for 'Yes' and 'No'.

Google search (knowledge graph)

A screenshot of a Google search result for a person. On the left, there's a sidebar with 'One World: Together At Home' and a 'Feedback' button. The main area shows a 'Available on' section with links to YouTube, Spotify, Google Play Music, and Deezer. Below this is a detailed card for Jacky Cheung Hok-yau, listing his birth (July 10, 1961), height (1.76 m), spouse (May Lo), and albums. A 'Songs' section is also present.

Google search

A screenshot of a Google search result for 'Agraphia'. It features two featured snippets: one from Wikipedia and another from a document. Below the snippets is a 'People also ask' section with a 'Feedback' button highlighted with a blue border.

Youtube music

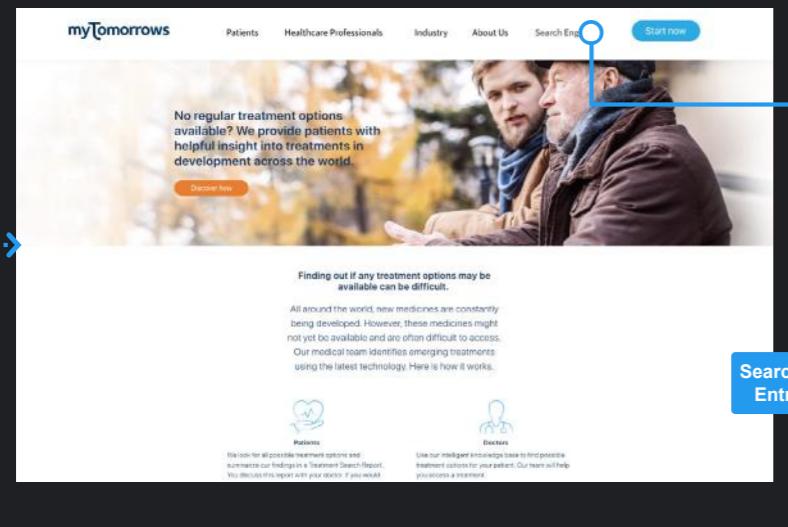
A screenshot of a YouTube music player. It shows a song titled 'High Alice' by Jenny Hval from the album 'The Practice of Love' (2019). The player includes standard controls for volume, play/pause, and seek.

Youtube

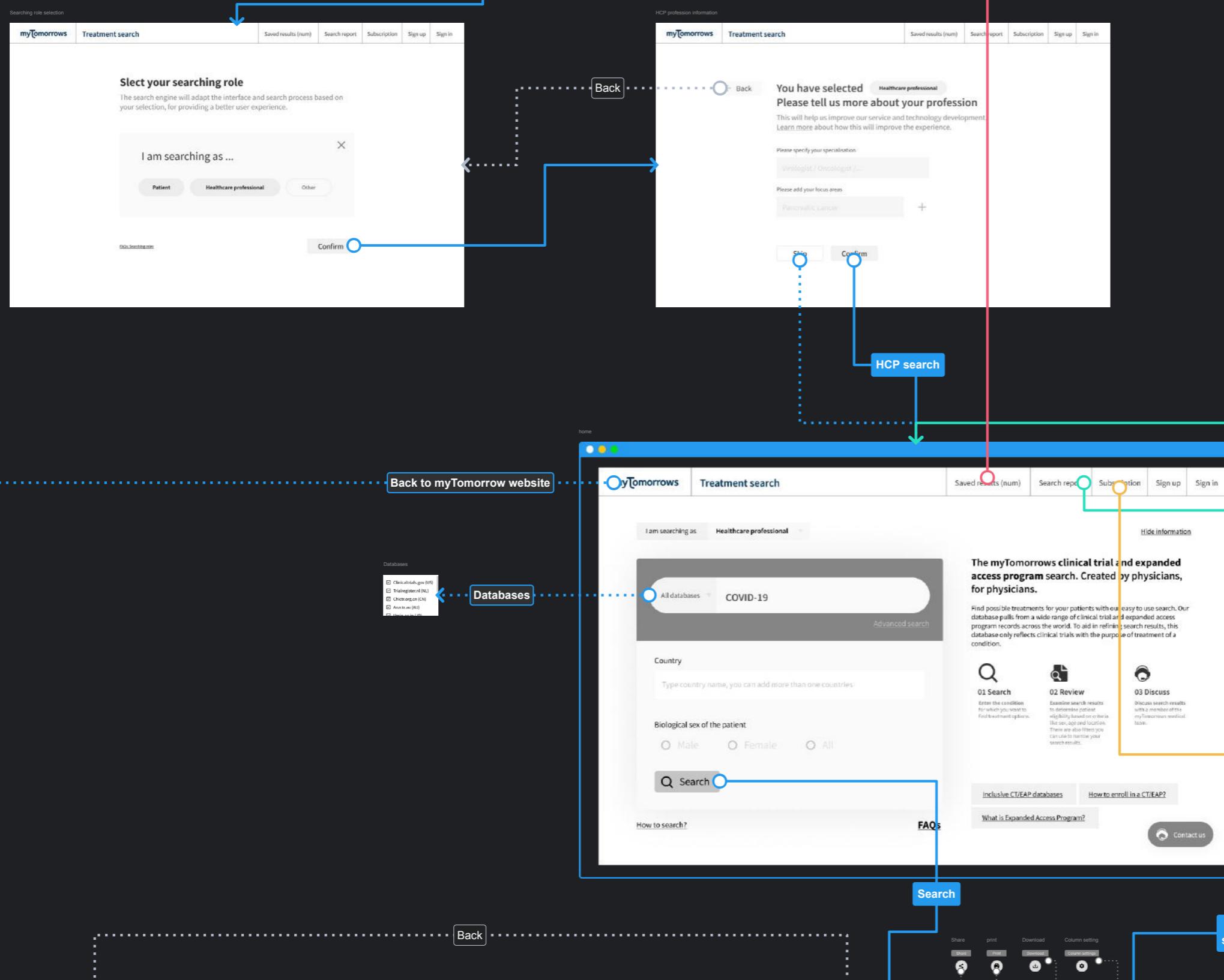
A screenshot of a YouTube channel page. It shows a video thumbnail for 'Mozart - Classical Music for Brain Power' and another for 'Mix - 王若琳 Joanna Wang - 有你的快樂'. The channel has 1M views and was created 1 year ago. A feedback menu is open at the bottom, with options 'Not interested' and 'Don't recommend channel' highlighted with a blue border.

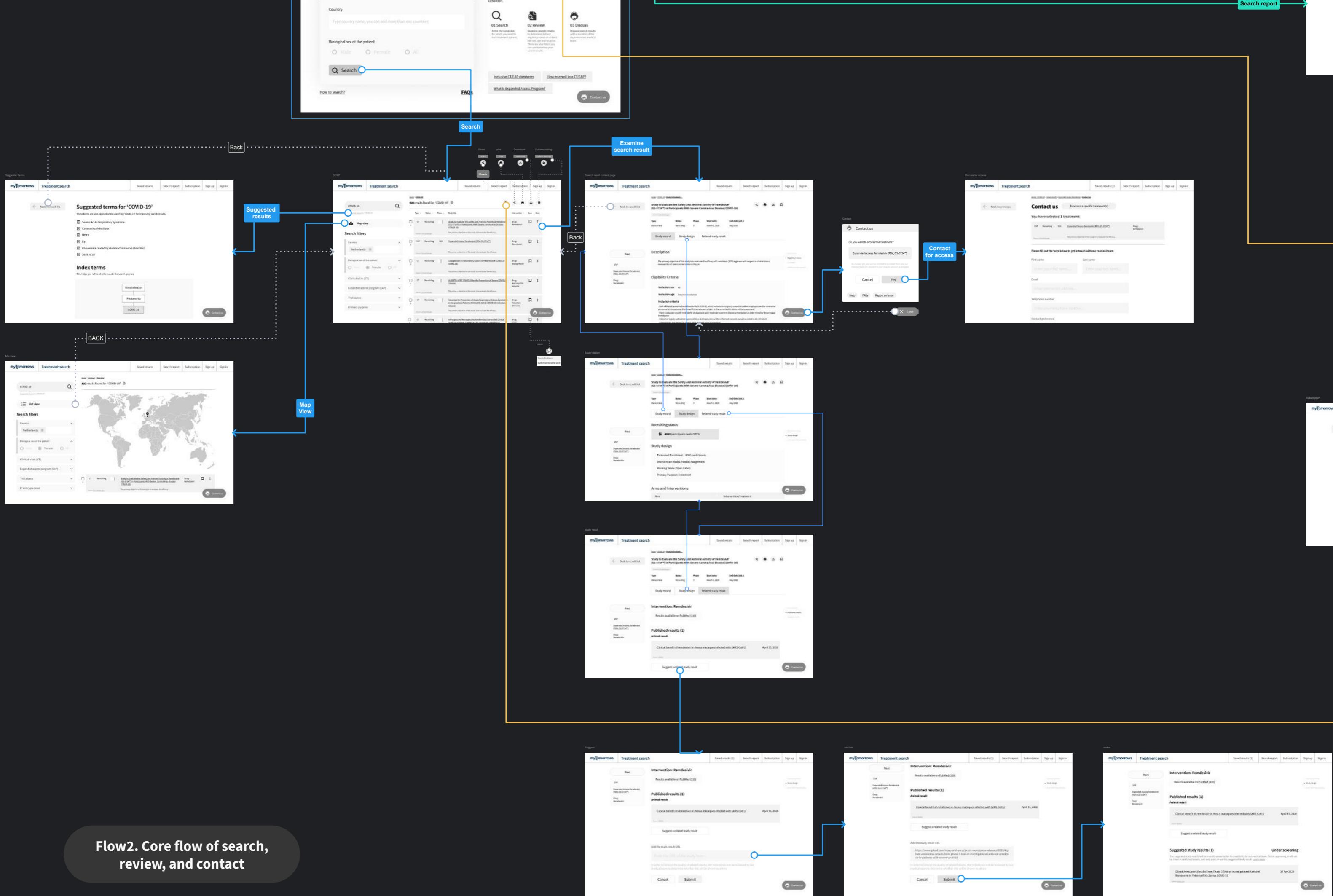
A screenshot of a YouTube video player during a performance of 'LIVE AID 1985'. A survey overlay is visible, asking 'Overall, how satisfied are you with this playlist?'. It includes a scale from 'Very dissatisfied' to 'Very satisfied' with corresponding smiley face icons. The video player shows the 'RIGHT SPEAKER' and 'LIVE AID 1985' text.

Specific feedback collection - Attached



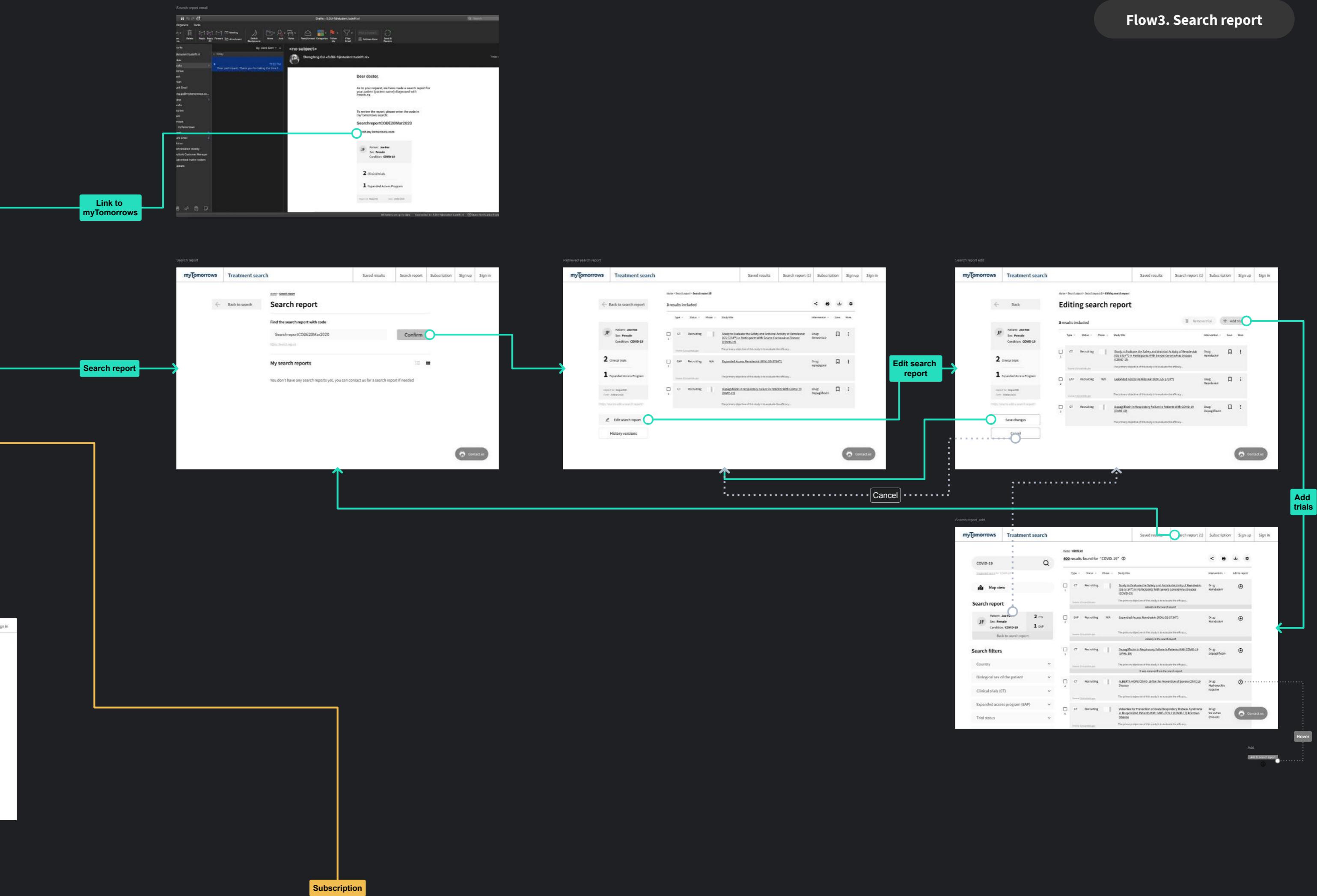
Flow1. myTomorrows search to search home



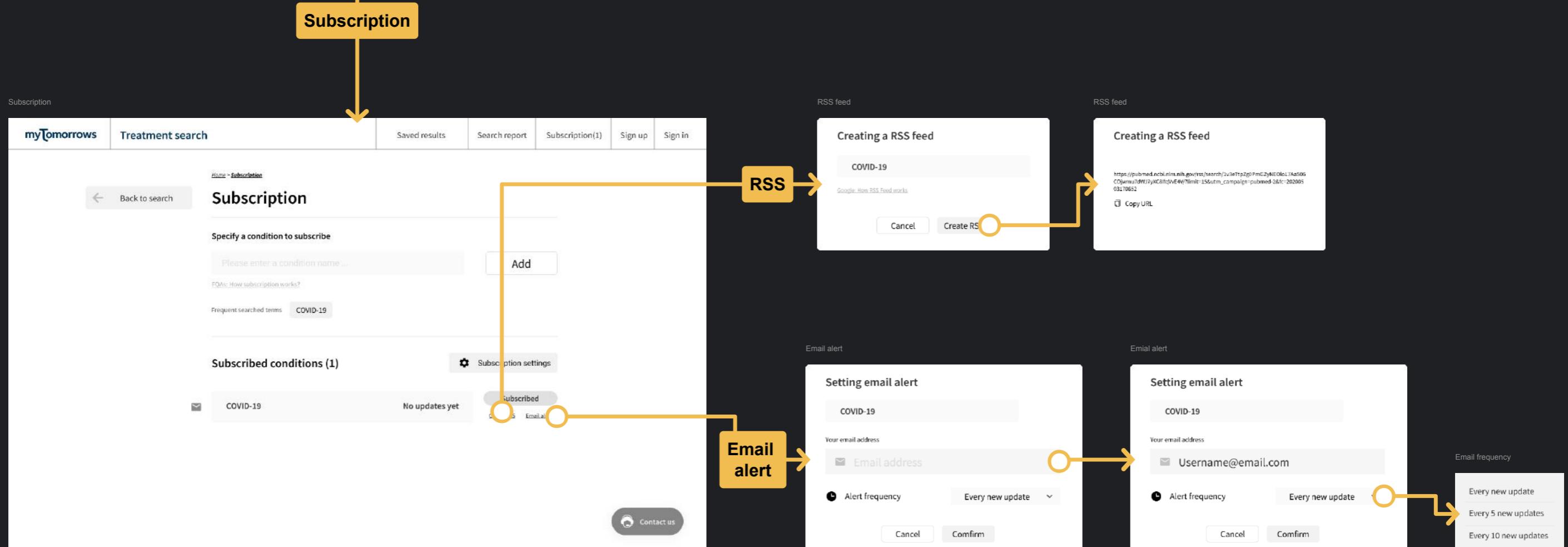


Flow2. Core flow of search, review, and contact

Flow3. Search report



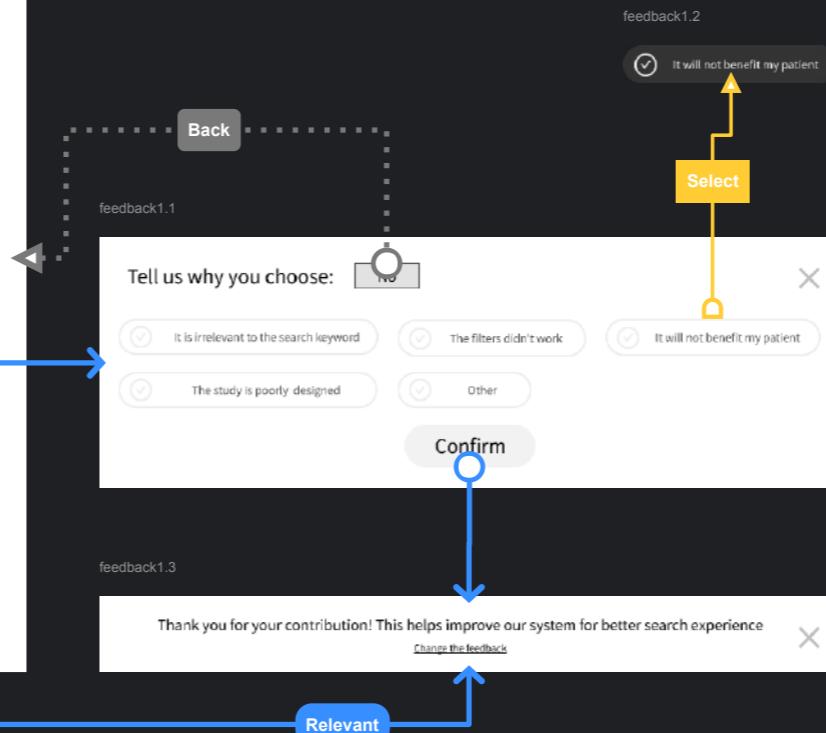
Flow4. Subscription



feedback1

The screenshot shows a clinical trial record for a study evaluating Remdesivir (GS-5734™) in COVID-19 patients. The trial is recruiting, Phase 3, starting March 6, 2020, and ending May 2020. The primary objective is to evaluate the efficacy of different remdesivir regimens. The page includes sections for Description, Eligibility Criteria, Inclusion criteria, and a feedback section at the bottom.

Feedback 1 - bottom popup



feedback2

The screenshot shows a study record for a trial with Trial ID NCT03469609, sponsored by Hannover Medical School and German Research Foundation. The page includes sections for Additional information, a feedback section at the bottom, and a footer with social media links.

Feedback 2 - bottom content

This screenshot shows the same study record as the previous one, but with a 'feedback2.1' overlay. The overlay contains a 'Tell us why you choose:' dropdown set to 'No', a list of reasons (Irrelevant, The filters didn't work, It will not benefit my patient, The study is poorly designed, Other), a 'Confirm' button, and a 'Thank you for your contribution!' message. The rest of the page content is visible below the overlay.

This screenshot shows the same study record as the previous ones, but with a 'feedback2.2' overlay. The overlay contains a 'Thank you for your contribution! This helps improve our system for better search experience' message and a 'Change the feedback' link. The rest of the page content is visible below the overlay.

feedback3

myTomorrows Treatment search

Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)

Type: Clinical trial Status: Recruiting Phase: 3 Start date: March 6, 2020 End date (est.): May 2020

Description

The primary objective of this study is to evaluate the efficacy of 2 remdesivir (RDV) regimens with respect to clinical status assessed by a 7-point ordinal scale on Day 14.

EAP: Expanded Access Remdesivir (RDV; GS-5734™)

Drug: Remdesivir

Eligibility Criteria

- Inclusion sex: All
- Inclusion age: Between 0 and older
- Inclusion criteria:
 - DoD-affiliated personnel as defined in DoDI 6200.02, which includes emergency-essential civilian employees and/or contractor personnel accompanying the Armed Forces who are subject to the same health risk as military personnel
 - Have a laboratory-confirmed COVID-19 diagnosis with moderate to severe disease presentation as determined by the principal investigator
 - Patient or legally authorized representative (LAR) provides written informed consent, except as noted in 21 CFR 50.23

Feedback

Relevant

Irrelevant

Not sure

Tell us why?

It is irrelevant to the search keyword
The filters didn't work
It is not useful for my patient
Other

Confirm

Irrelevant

feedback3.1

feedback3.2

feedback3.3

feedback3.4

feedback3.1

feedback3.2

feedback3.3

feedback3.4

Feedback 3 - Action + Selection

feedback4

myTomorrows Treatment search

Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)

Type: Clinical trial Status: Recruiting Phase: 3 Start date: March 6, 2020 End date (est.): May 2020

Description

The primary objective of this study is to evaluate the efficacy of 2 remdesivir (RDV) regimens with respect to clinical status assessed by a 7-point ordinal scale on Day 14.

EAP: Expanded Access Remdesivir (RDV; GS-5734™)

Drug: Remdesivir

Eligibility Criteria

- Inclusion sex: All
- Inclusion age: Between 0 and older
- Inclusion criteria:
 - DoD-affiliated personnel as defined in DoDI 6200.02, which includes emergency-essential civilian employees and/or contractor personnel accompanying the Armed Forces who are subject to the same health risk as military personnel
 - Have a laboratory-confirmed COVID-19 diagnosis with moderate to severe disease presentation as determined by the principal investigator
 - Patient or legally authorized representative (LAR) provides written informed consent, except as noted in 21 CFR 50.23

Feedback

Useful

Useless

Not sure

Negative

Tell us why?

Please tell us why?

Select element that determines your judgement

Confirm

Negative

feedback4.1

feedback4.2

feedback4.3

feedback4.4

feedback4.1

feedback4.2

feedback4.3

feedback4.4

Feedback 4 - Action + ‘Pointing out’

feedback5

This screenshot shows the treatment search results for COVID-19. A specific study entry is highlighted: "Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)". The study is an "Expanded Access Program" available at N/A status, starting on February 22, 2020, and ending on July 1, 2020. The study record tab is selected. On the right, there's a "Description" section with a "Next" button, an "Eligibility Criteria" section, and a "Contact us" button.

This screenshot shows the treatment search results for COVID-19. A specific study entry is highlighted: "Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)". The study is an "Expanded Access Program" available at N/A status, starting on February 22, 2020, and ending on July 1, 2020. The study record tab is selected. On the right, there's a "Description" section with a "Previous" button, an "Eligibility Criteria" section, and a "Contact us" button. A blue arrow points from the "Review the next" button in the first screenshot to the "Next" button in this screenshot.

Feedback 5 - Pop-up comparison

Feedback 6 - Surrogate message

This screenshot shows the treatment search results for COVID-19. A study entry for "Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)" is highlighted. The study is an "Expanded Access Program" recruiting at N/A status, starting on February 22, 2020, and ending on July 1, 2020. The study record tab is selected. On the right, there's a "Description" section with a "Next" button, an "Eligibility Criteria" section, and a "Contact us" button. A blue arrow points from the "Review the next" button in the previous screenshot to the "Next" button in this screenshot.

This screenshot shows the treatment search results for COVID-19. A study entry for "Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)" is highlighted. The study is an "Expanded Access Program" recruiting at N/A status, starting on February 22, 2020, and ending on July 1, 2020. The study record tab is selected. On the right, there's a "Description" section with a "CT" button, an "Eligibility Criteria" section, and a "Contact us" button. A blue arrow points from the "Review the next" button in the previous screenshot to the "CT" button in this screenshot.

This screenshot shows the treatment search results for COVID-19. A study entry for "Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734™) in Participants With Severe Coronavirus Disease (COVID-19)" is highlighted. The study is an "Expanded Access Program" recruiting at N/A status, starting on February 22, 2020, and ending on July 1, 2020. The study record tab is selected. On the right, there's a "Description" section with a "CT" button, an "Eligibility Criteria" section, and a "Contact us" button. A blue arrow points from the "Review the next" button in the previous screenshot to the "Positive" button in this screenshot. A feedback overlay at the top right says "Thank you for your contribution! This helps improve our system for better search experience. Change the feedback".

feedback7

Feedback Session

feedback7.1

Feedback 7 - Task session

feedback8.2

Promising bin (1)

Drop

feedback8.1

Drop

feedback8.3

Trash bin (1)

feedback8.4

Uncertain bin (1)

Feedback 8 - Feedback bins

Feedback 9 - Feedback bot

The image contains two screenshots of the myTomorrows Treatment search interface. The top screenshot shows a search results page for "COVID-19" with several study cards. A blue box highlights the first study card, and a red arrow points from the "Examine Result" button to this area. A blue line connects this to the bottom screenshot. The bottom screenshot shows a detailed view of the first study card, specifically the "Eligibility Criteria" section. A blue box highlights this section, and a red arrow points from the "Examine Result" button to it. A blue line connects this to the top screenshot.

feedback9

Examine Result

feedback9.1

Hover

feedback9.2

Lv. 5

feedback9.3

Examine Result

Treatment search survey

7 responses

[Publish analytics](#)

Your name

7 responses

elyn

Sum Yuet Leung

Yao-Jun Chen

Perry

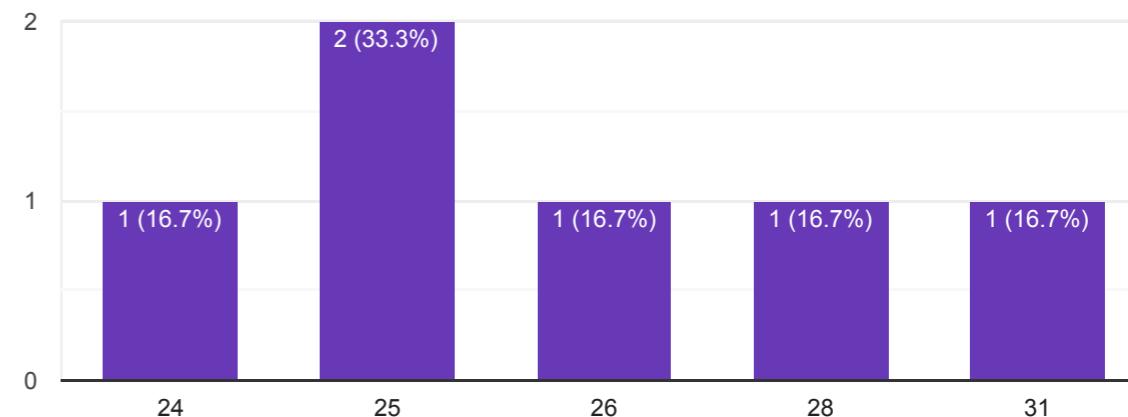
Angela

Maria

Junyao Li

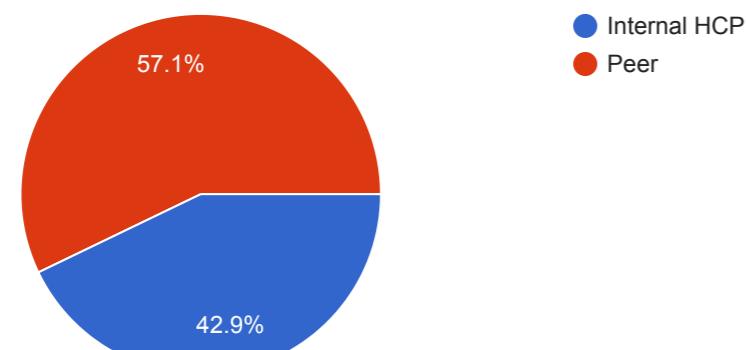
Age

6 responses

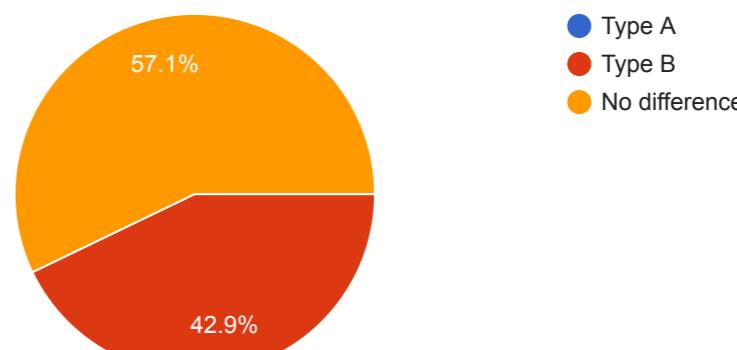


Testing role

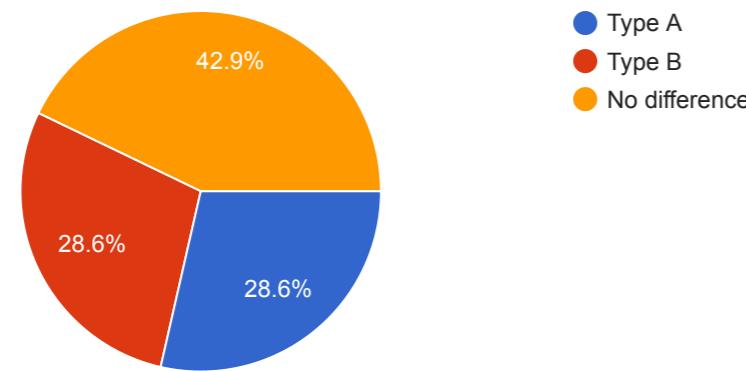
7 responses

**Type A or B provides a better experience for task 2. saving results?**

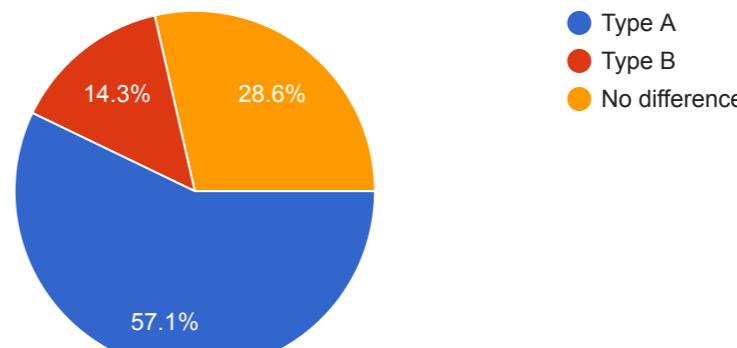
7 responses

**Session 1 experience****Type A or B provides a better experience for task 1. searching and reviewing?**

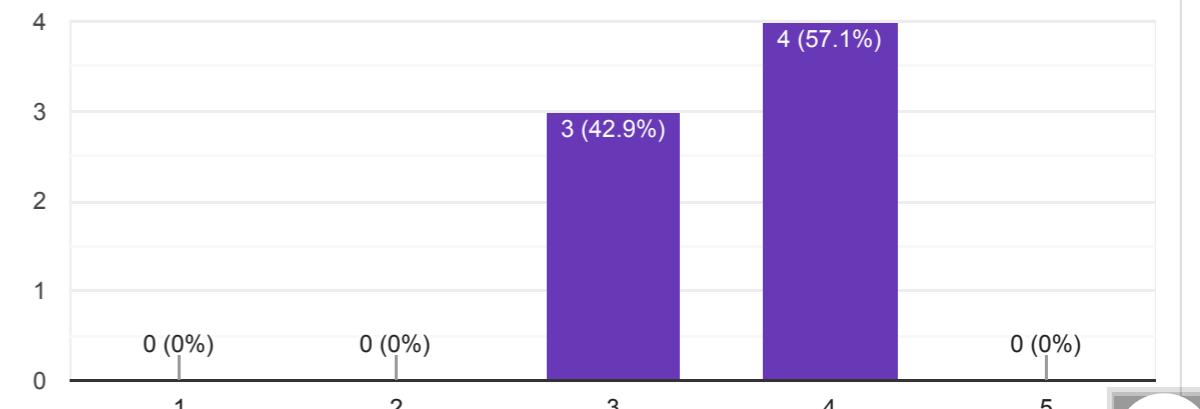
7 responses

**Type A or B provides a better experience for task 3. gaining access to a result?**

7 responses

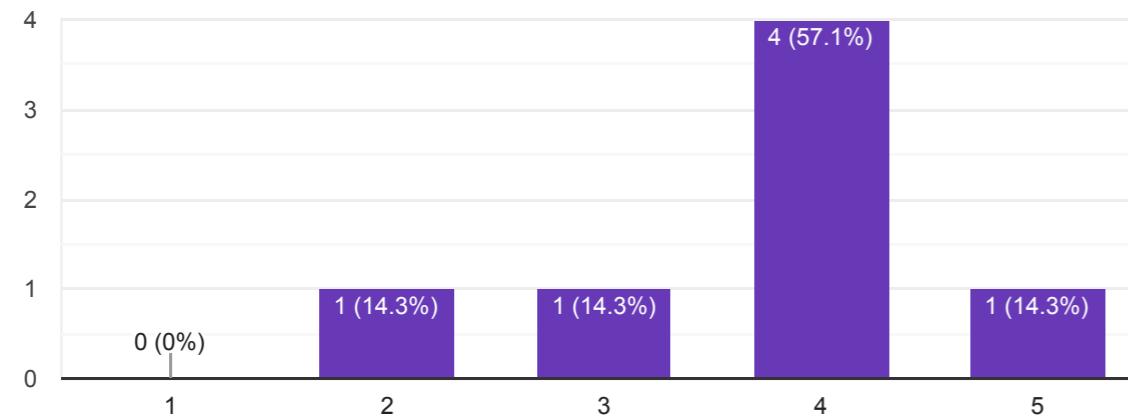
**The overall experience you have with Type A**

7 responses



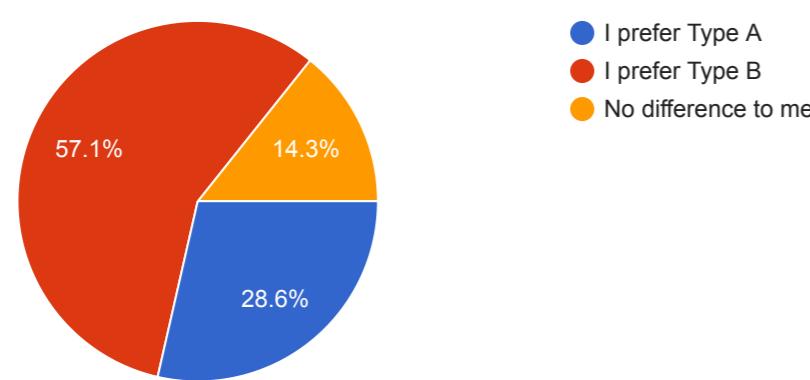
The overall experience you have with Type B

7 responses

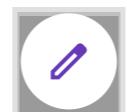


Would you prefer Type A or Type B overall?

7 responses

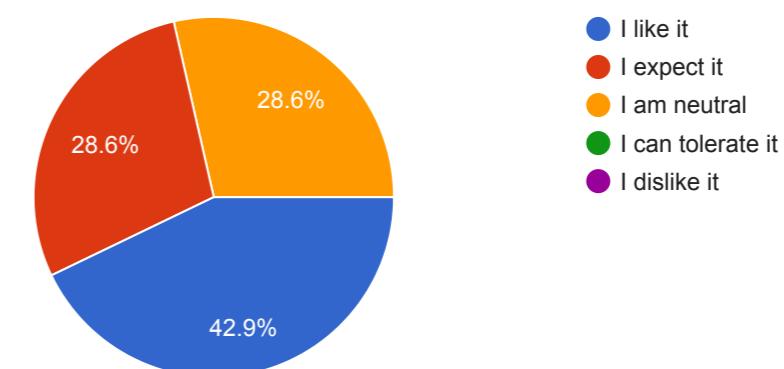


Session 2 experience



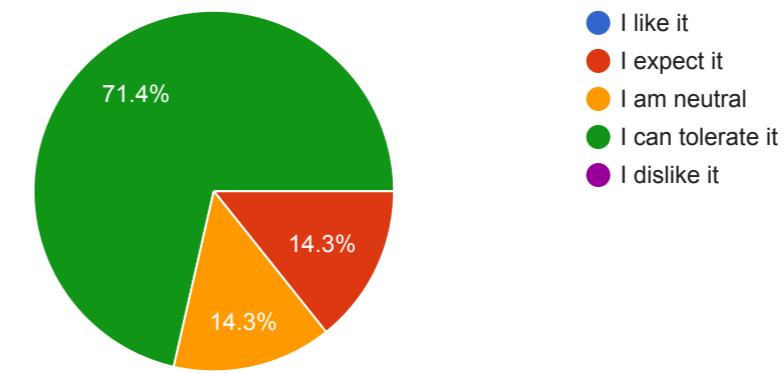
Having 'Subscription' feature

7 responses



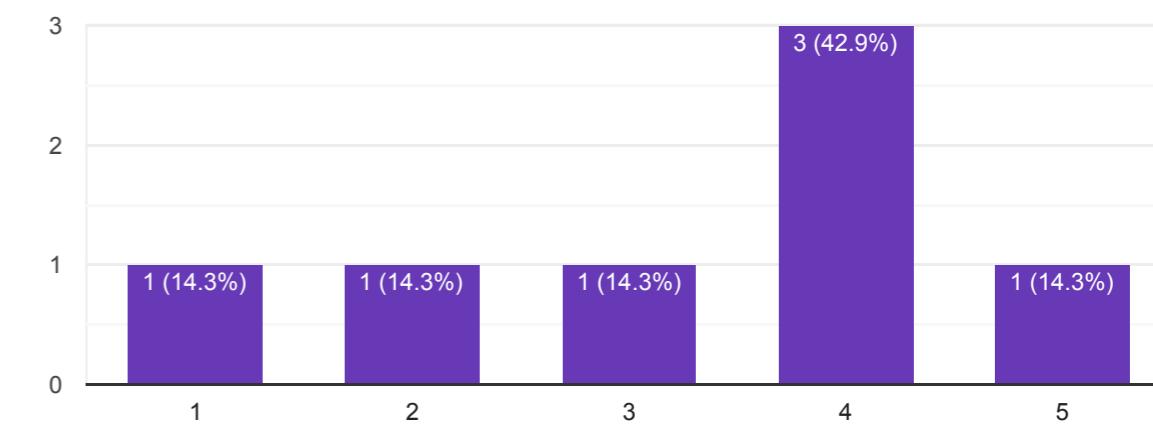
Not having 'Subscription' feature

7 responses



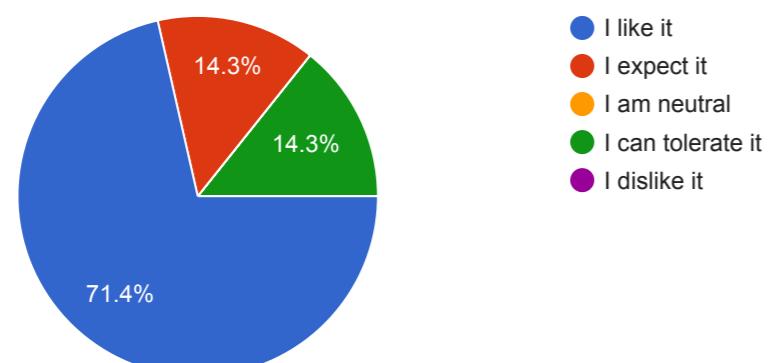
How was your experience interacting with 'Subscription' feature?

7 responses

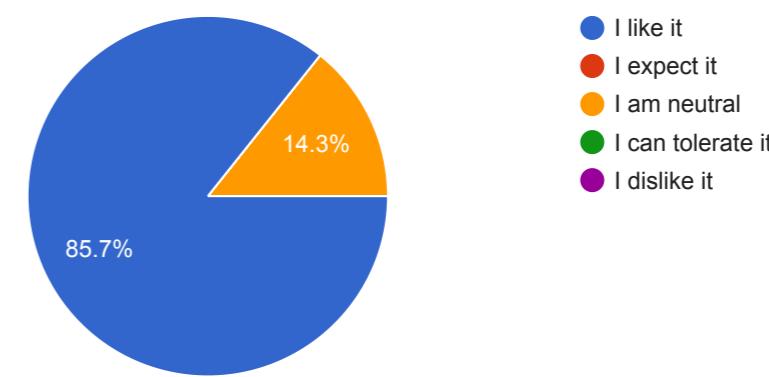


Having 'Search report' feature

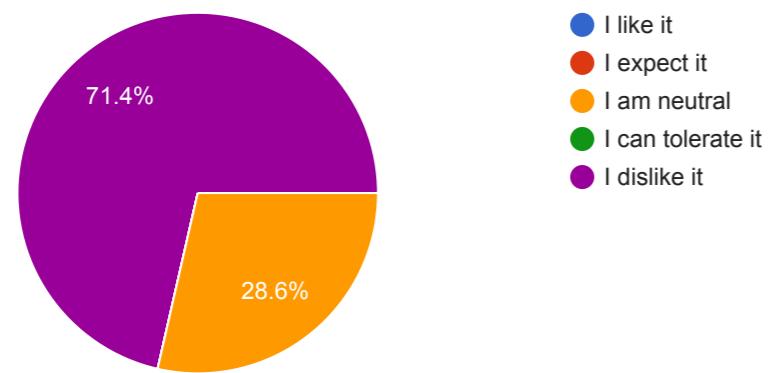
7 responses

**Having 'Folder' feature**

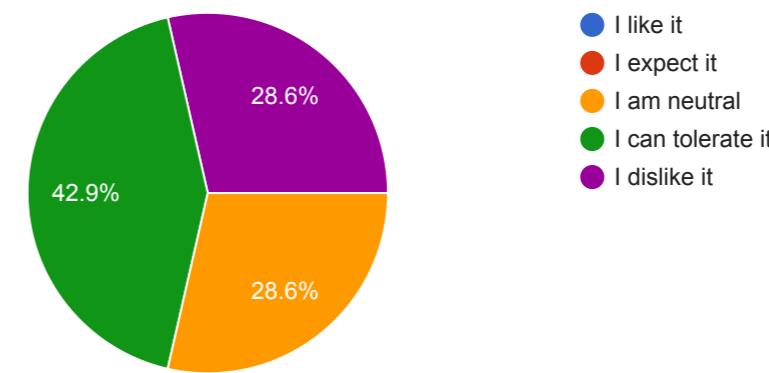
7 responses

**Not having 'Search report' feature**

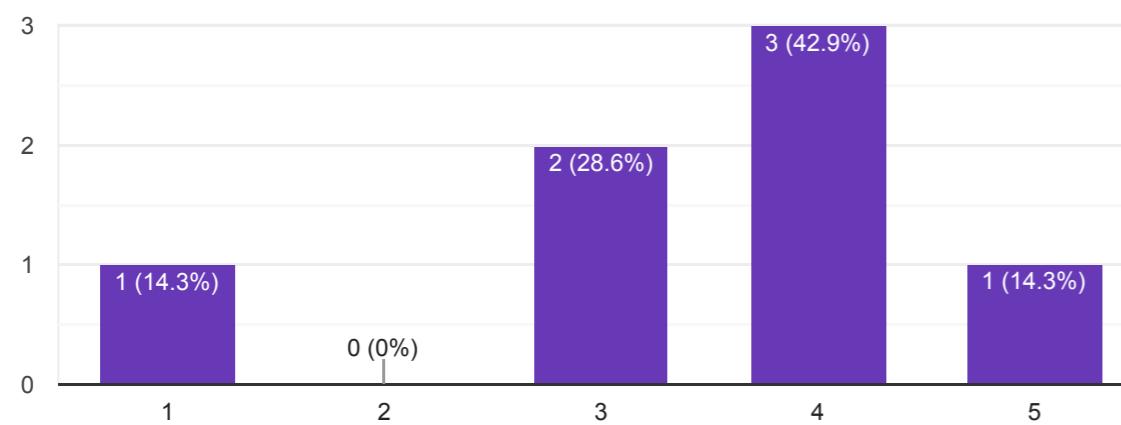
7 responses

**Not having 'Folder' feature**

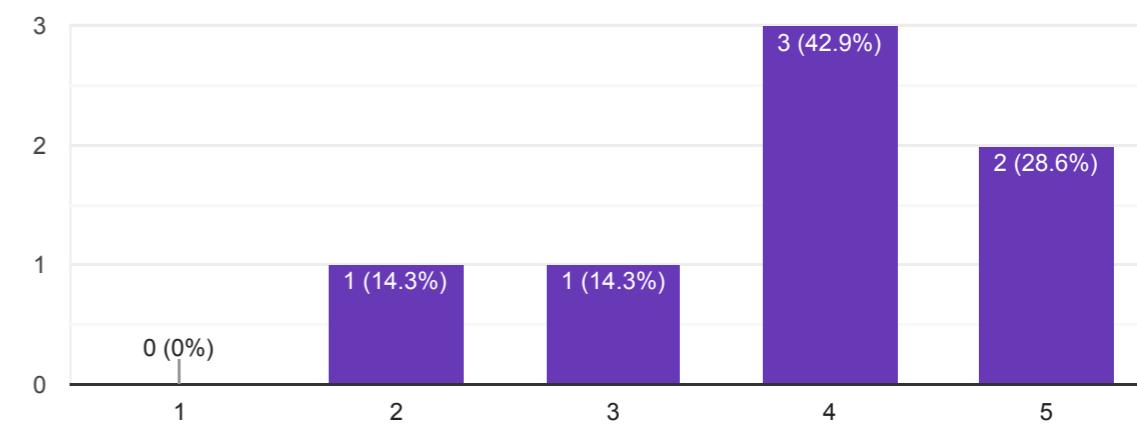
7 responses

**How was your experience interacting with the 'Search report' feature?**

7 responses

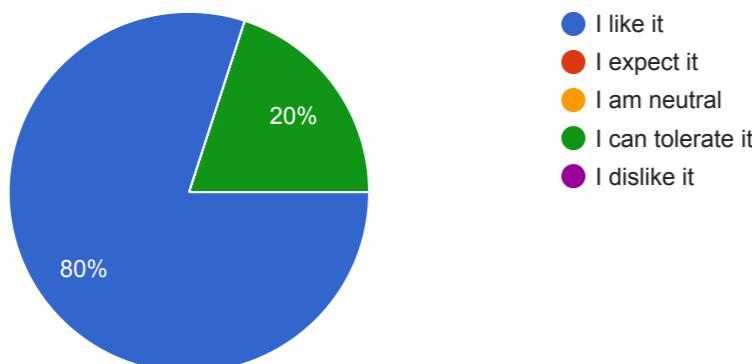
**How was your experience interacting with the 'Folder' feature?**

7 responses



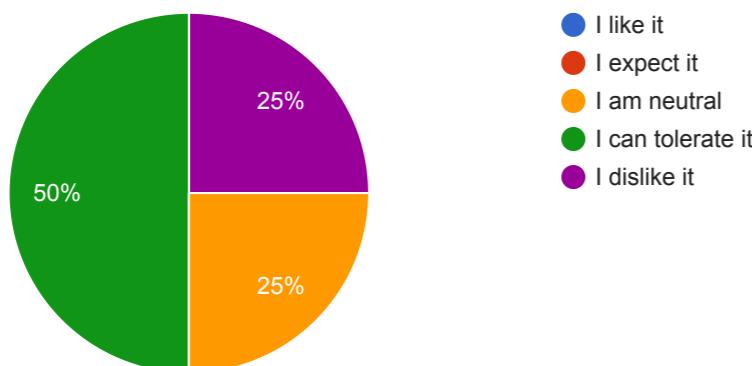
Having 'Suggest study result' feature

5 responses



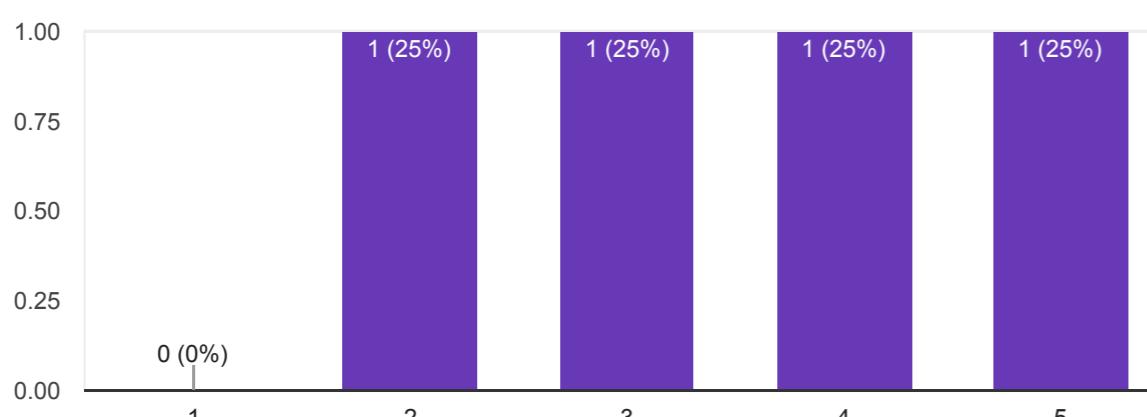
Not having 'Suggest study result' feature

4 responses



How was your experience interacting with 'Suggest study result' feature?

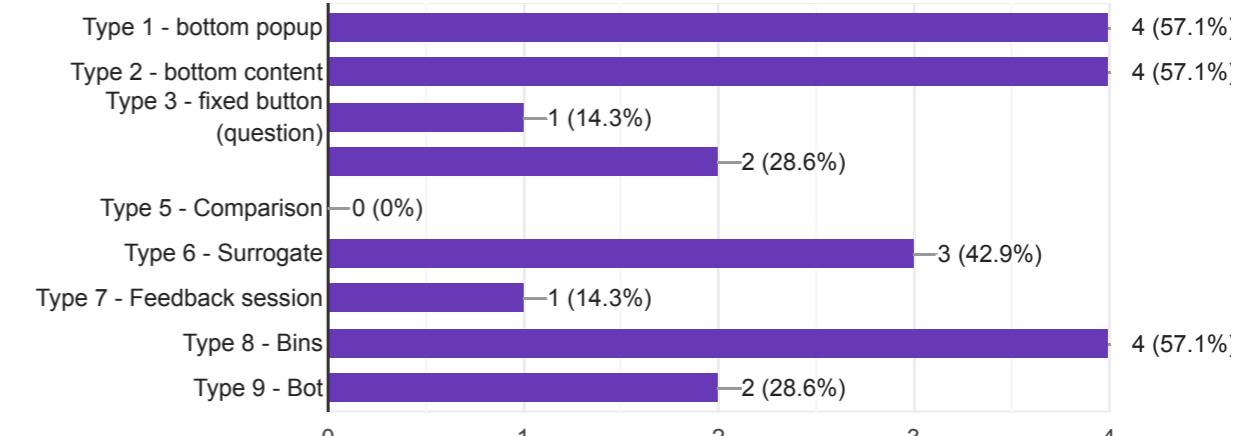
4 responses



Session 3 (Feedback)

Please pick top 3 feedback interactions you prefer the best

7 responses



Please pick the favorite one and specify the reasons

7 responses

9

type5

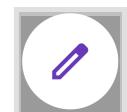
1+2 combination

6: most easy

Type 8

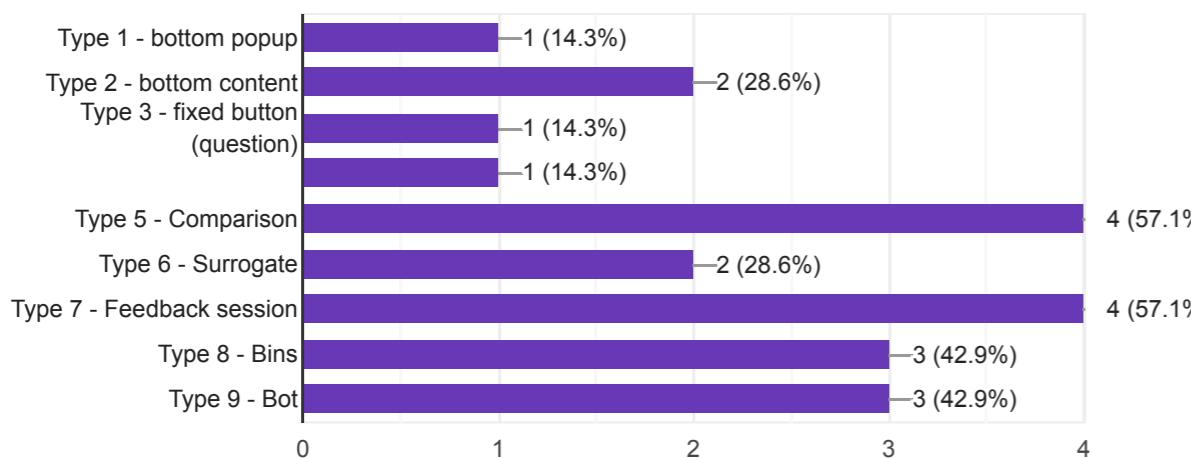
1

Type 1



Please pick 3 feedback interactions you dislike the most

7 responses



Please pick the most disliked one and specify the reasons

7 responses

7

type7

8

1: you have to click it away

Type 9

5

Type7

This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)

Google Forms



Treatment search test questionnaire

9 responses

[Publish analytics](#)

Your name

9 responses

Dominic van den Brock

Ciya Liang

Carlijn Jordaan

Anna Torrecaan

Rosanne Verwijs

Casper Rokx

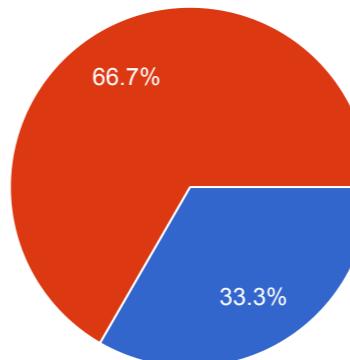
Bram Dierckx

Chenow Wang

Katie Hensley

Your age

9 responses

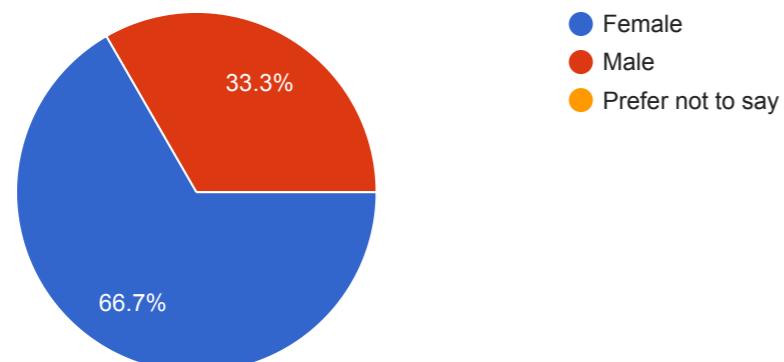


- 18 - 24
- 25 - 44
- 45 - 64
- > 65



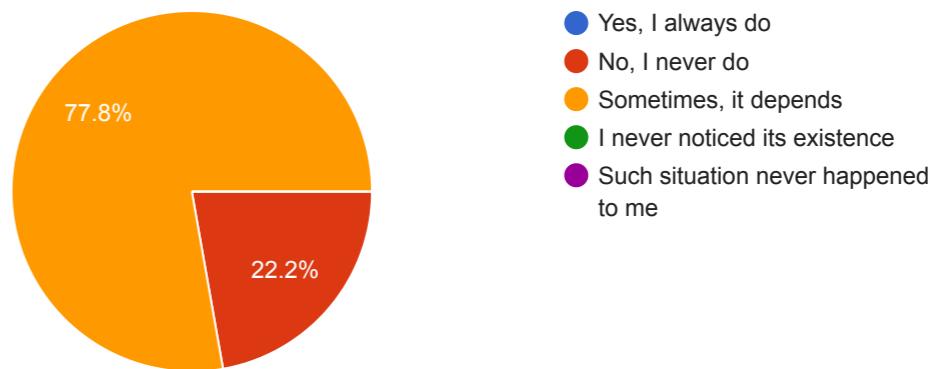
Gender

9 responses



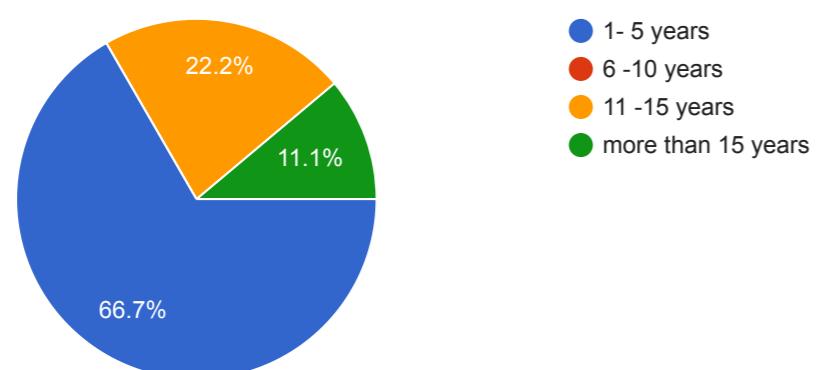
With your previous experience, do you provide feedback in such (similar) situations?

9 responses



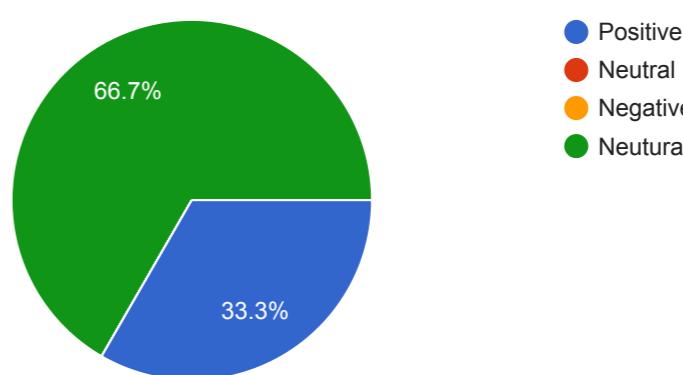
How many years have you become a health professional?

9 responses



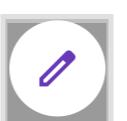
How do you feel providing feedback in such (similar) situations?

9 responses



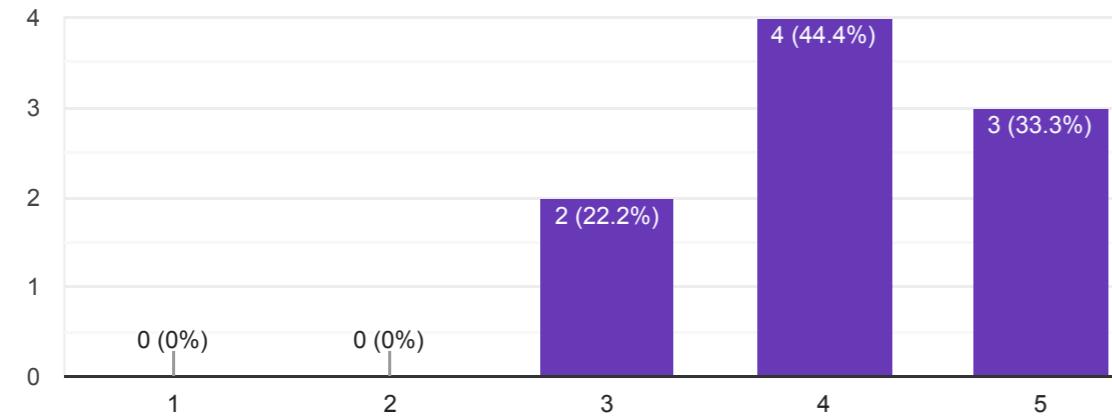
Your attitude towards feedback

Your experience with the treatment search



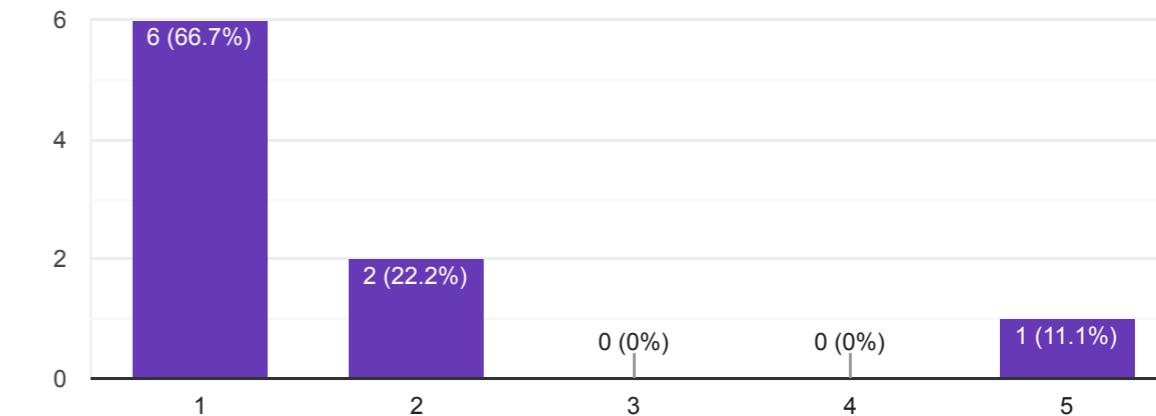
How was your overall experience with the treatment search?

9 responses



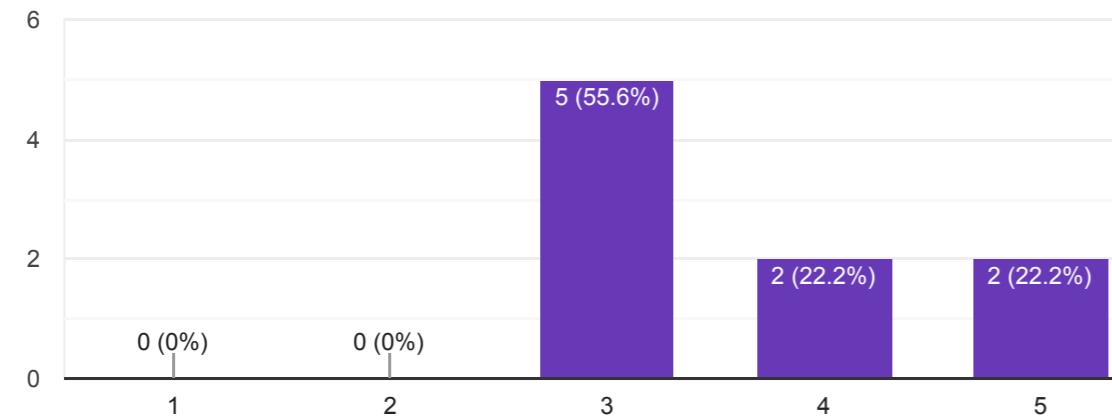
I found the system unnecessarily complex.

9 responses



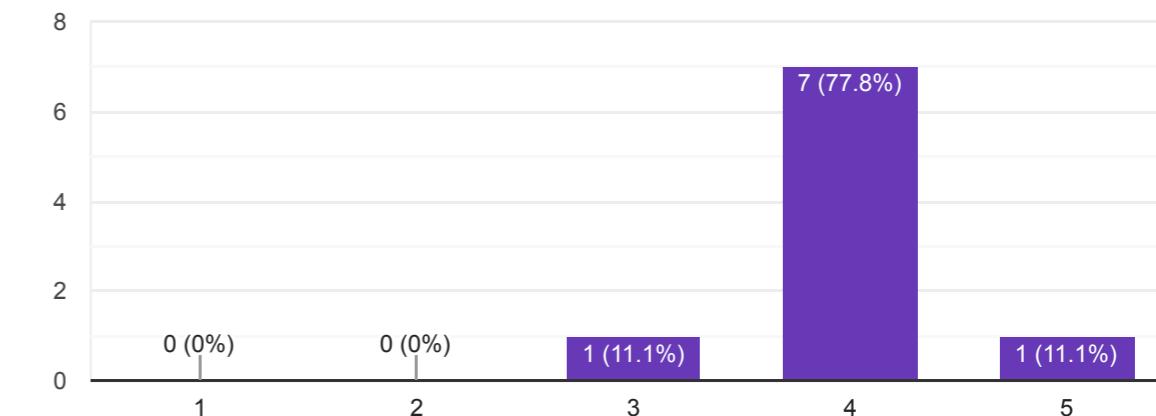
I think that I would like to use this system frequently.

9 responses



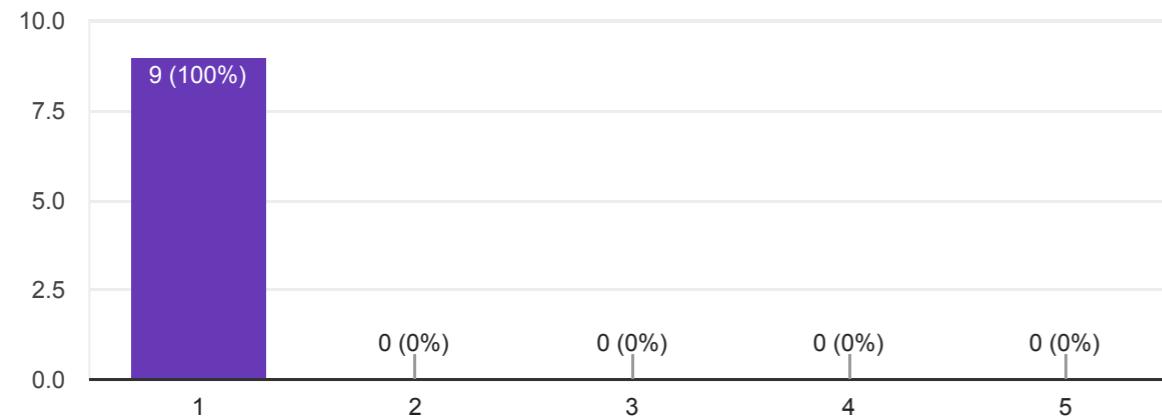
I thought the system was easy to use.

9 responses



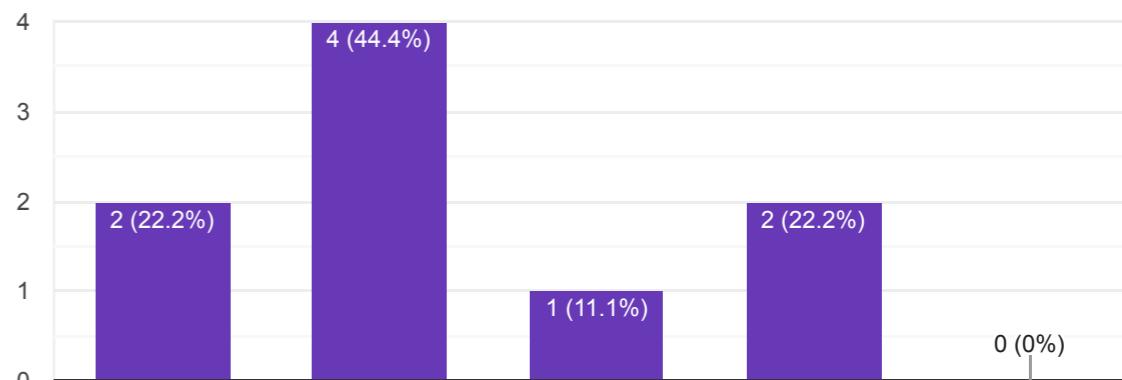
I think that I would need the support of a technical person to be able to use this system.

9 responses



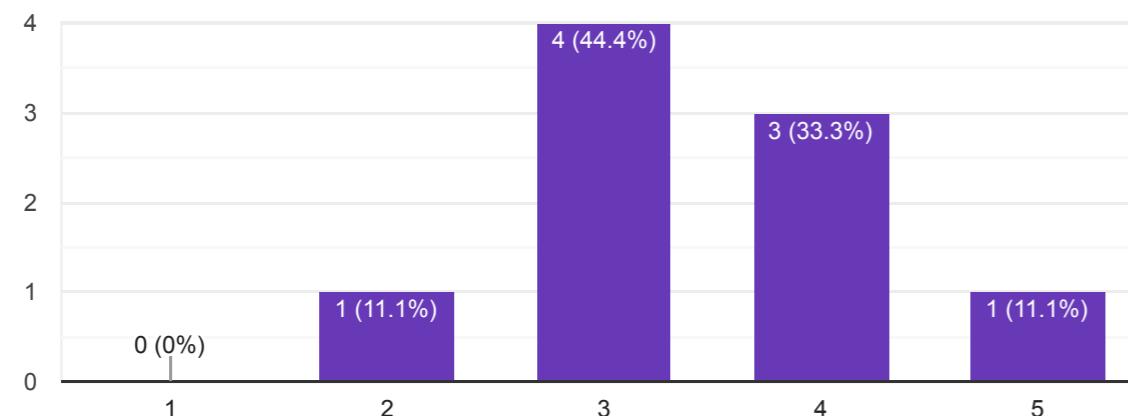
I thought there was too much inconsistency in this system.

9 responses



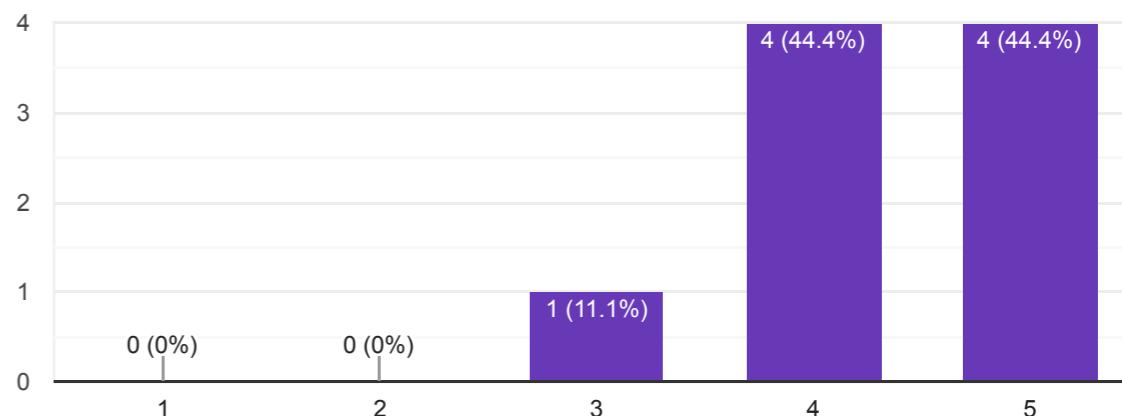
I found the various functions in this system were well integrated.

9 responses



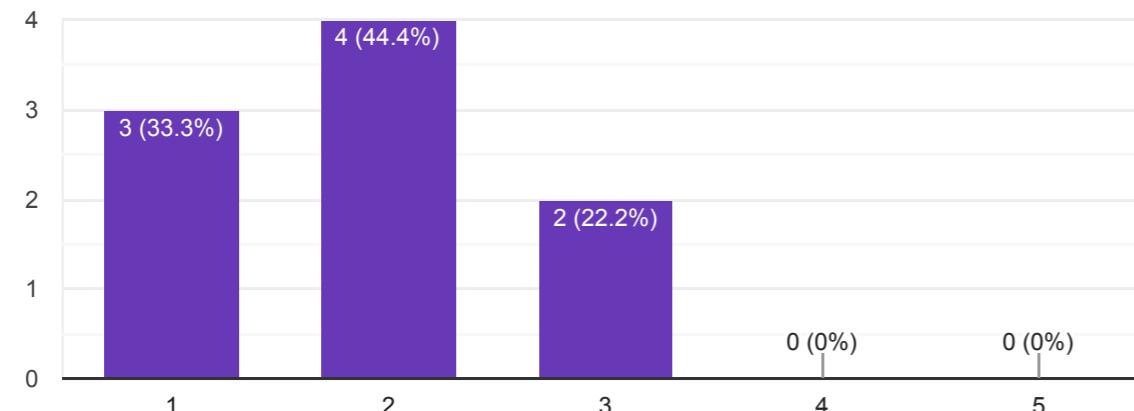
I would imagine that most people would learn to use this system very quickly.

9 responses



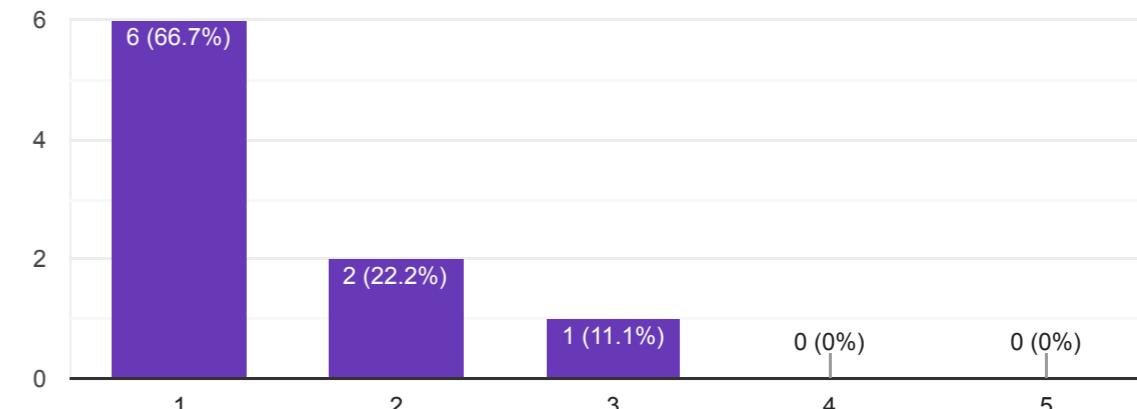
I found the system very cumbersome to use.

9 responses



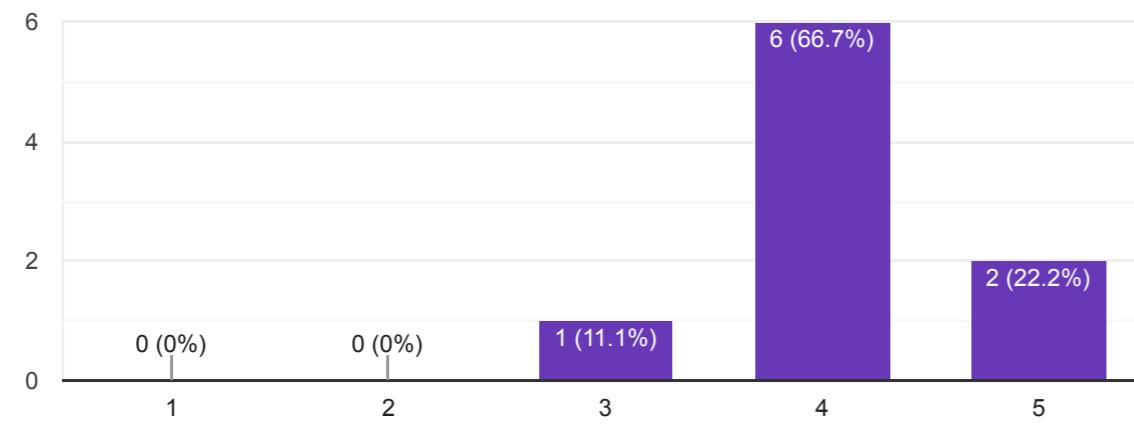
I needed to learn a lot of things before I could get going with this system.

9 responses



I felt very confident using the system.

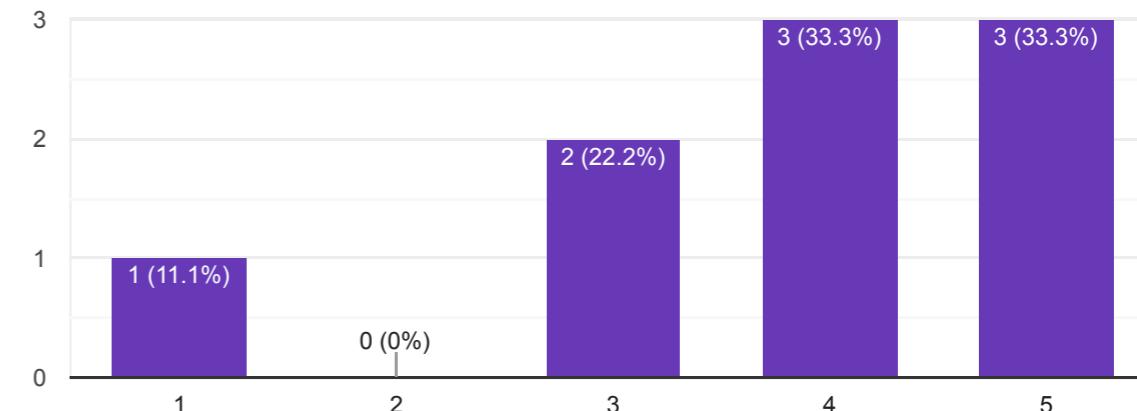
9 responses



Feedback collection experience

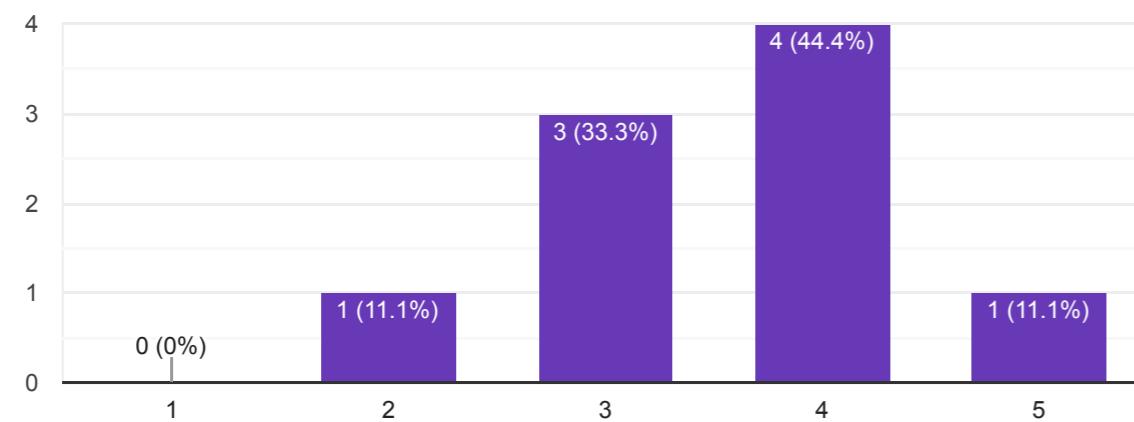
Experience with the 1st concept_Pre-screener

9 responses



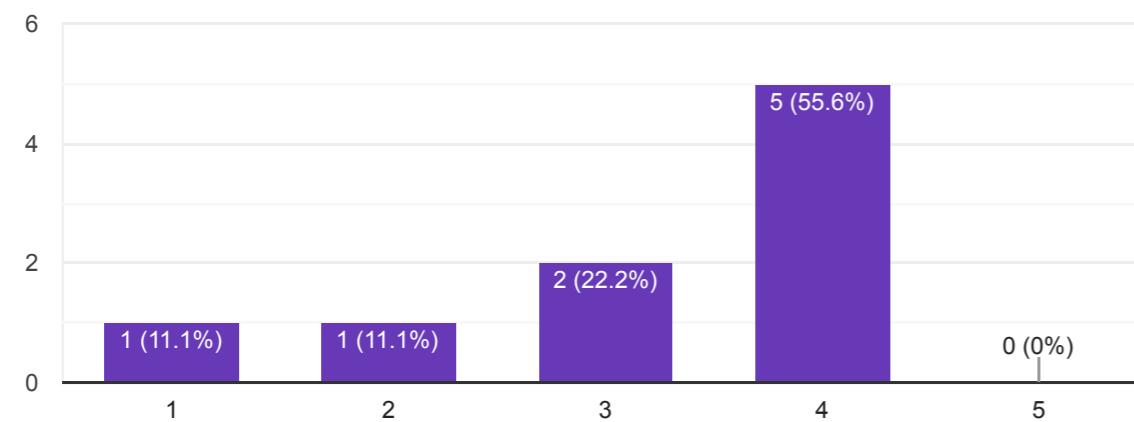
Experience with the 2nd concept_Reminder

9 responses



Experience with the 3rd concept_Hearty

9 responses

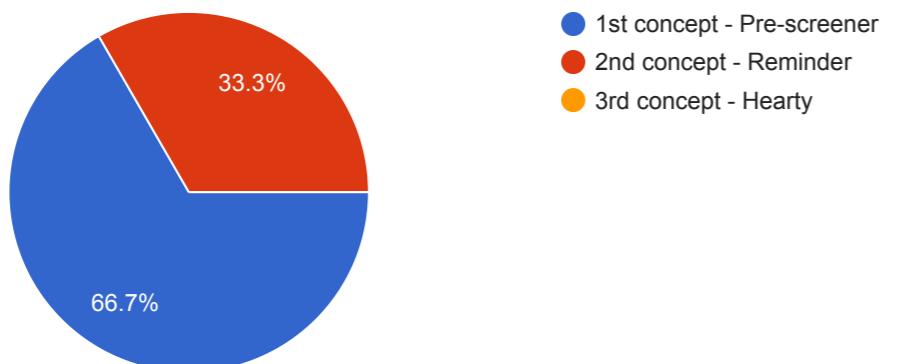


Home page	<ul style="list-style-type: none"> • Cleaner interface to reduce visual burdens • Swap Treatment search and myT LOGO
Search result page	<ul style="list-style-type: none"> • Expanded filters as default • Add Icons to navigation bar options (better navigation) • Enable hover to preview
Content page	<ul style="list-style-type: none"> • Strengthen explanation and guidance on gaining access • Merge study record and study design into one page • Change the string 'related study result' • Provide data visualization of study results
General	<ul style="list-style-type: none"> • Optimize checkbox interactions • Optimize visuals to increase readability and information hierarchy
Subscription	<ul style="list-style-type: none"> • Offer more explanation
Search report	<ul style="list-style-type: none"> • Add more information on report generation to increase the feeling of a report • Enhance the explanation of what is search report • Increase guidance and explanation of the next steps to take • Remove 'edit report' and 'saves changes', and allow direct edit of search report
myFolders	<ul style="list-style-type: none"> • Results in folders under Saved results are duplicated from the all saved results, instead of moved, to avoid a blank saved result page.

Table 19. Table of UX improvements by key pages and features.

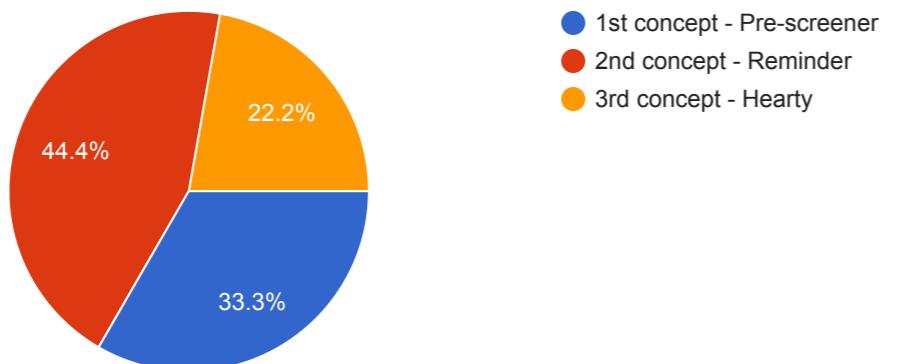
Which is the favorite concept of collecting feedback?

9 responses



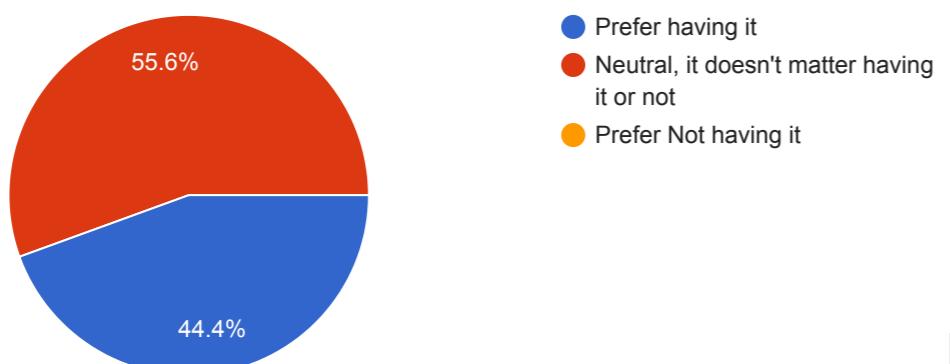
Which feedback collection concept motivates you the most to provide feedback?

9 responses



Consider the favorite one you have chosen, would you prefer having it in the treatment search or not?

9 responses



DESIGN
FOR OUR
future

3961

TU Delft

IDE Master Graduation

Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

! USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT

Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

STUDENT DATA & MASTER PROGRAMME

Save this form according the format "IDE Master Graduation Project Brief_familyname_firstname_studentnumber_dd-mm-yyyy". Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

family name Gu
initials S.G given name Shengfeng
student number 4813456
street & no. Prof. Schermerhornstraat 103-3
zipcode & city 2628 PZ
country the Netherlands
phone 0626529031
email goooleon@gmail.com

Your master programme (only select the options that apply to you):

IDE master(s): IPD Dfl SPD

2nd non-IDE master: _____

individual programme: _____ (give date of approval)

honours programme: Honours Programme Master

specialisation / annotation: Medisign

Tech. in Sustainable Design

Entrepreneurship

Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v..

Second mentor only applies in case the assignment is hosted by an external organisation.

Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.

SUPERVISORY TEAM **

Fill in the required data for the supervisory team members. Please check the instructions on the right!

** chair Alessandro Bozzon dept. / section: Design Engineering
** mentor Derek Lomas dept. / section: Industrial Design
2nd mentor Zoltán Szlávik
organisation: myTomorrows
city: Amsterdam country: the Netherlands

comments
(optional)

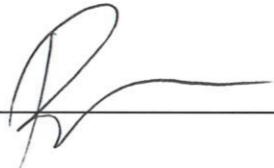
APPROVAL PROJECT BRIEF

To be filled in by the chair of the supervisory team.

chair Alessandro Bozzon

date 21-03-2020

signature



CHECK STUDY PROGRESS

To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair. The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: 30 EC

Of which, taking the conditional requirements into account, can be part of the exam programme 30 EC

List of electives obtained before the third semester without approval of the BoE

 YES all 1st year master courses passed

 NO missing 1st year master courses are:

name _____

date 10-3-2020

signature CB

FORMAL APPROVAL GRADUATION PROJECT

To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked **. Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks ?
- Does the composition of the supervisory team comply with the regulations and fit the assignment ?

Content: APPROVED NOT APPROVEDProcedure: APPROVED NOT APPROVED

comments

name Manon Borgstijn

date 02-04-2020

signature MB

Towards a clear and effective interactive search experience

project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 17 - 02 - 2020

15 - 07 - 2020 end date

INTRODUCTION **

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

Artificial Intelligence (AI) is expected to play a transformative role in health and human wellbeing. Search and information retrieval technologies already play a major role in healthcare research and practice (e.g. by supporting continuing education and systematic reviews, Byron et al., 2012).

What does the future hold for intelligent search in healthcare? myTomorrows, a pharma-tech company (figure 1) based in Amsterdam, uses AI and supporting information services to improve the access to and recruitment of novel treatments in development (e.g.Clinical trials, Expanded Access Programs, Off-label Use, etc.). By engaging more potential participants and matching them to clinical trials, myTomorrows can provide access to life-saving new medicines for patients and advance the pace of medical discovery. However, their current search and recommendation system (figure 2) poses several user experience challenges that may interrupt the workflows of internal teams, slow down the pace of match-making and limit their impact.

How might Human-centered design methods support the design of AI systems within the field of health and well-being? This project will probe the problem myTomorrows search is exposed to through the lens of design and use of Human-centered design methods. By putting together the multidisciplinary knowledge and future visions, this research aims at improving the search experience and unleashing the full potential of myTomorrows AI recommendation system, thus, achieving a clear and effective search experience.

space available for images / figures on next page

introduction (continued): space for images

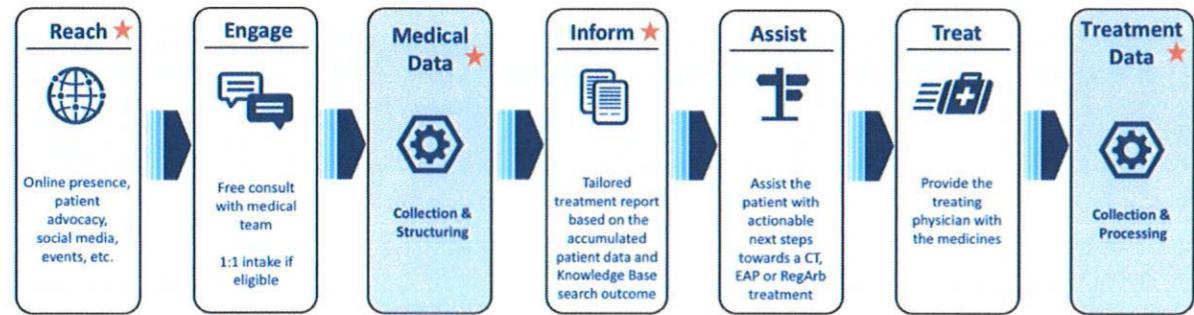


image / figure 1: Overview of myTomorrows service by step

myTomorrows treatment search

Search for possible treatment options in our Clinical Trials and Expanded Access Programs database

Our search pulls from a wide range of clinical trial and expanded access program databases from all over the world. Searching can become an overwhelming experience due to the complex medical landscape. To simplify the search process, we have preselected search filters that are most relevant for patients seeking possible treatment options

Searching as: Patient, Healthcare professional

Search criteria: Condition (Type and select disease), Country (Type the country, Note: Leave empty to search worldwide), Sex of patient (Male, Female, All)

Intake Call: This short call with our medical team lets us understand the specifics of your situation.

Search Report: You will receive a tailor made report listing treatment options that may be available to you.

Access: You can select a treatment together with your physician. myTomorrows can help your physician get access to this treatment.

Contact us

image / figure 2: myTomorrows search system (external version)

PROBLEM DEFINITION **

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

Search, in this specific context of myTomorrows, is to match patients to relevant treatment(s) with the support of healthcare professionals. Currently, however, two main problems are slowing down the process.

First, the unmet urge to improve the current search UX of guiding users throughout the system. a search does not guarantee the retrieval of relevant information, even at the very beginning. Forming search queries or keywords can be rather frustrating, and filtering relevant results from irrelevant ones may require detours or huge time investment. Search, as a labor-intensive process manifesting complex behavior patterns (search patterns, 2010), it is intriguing to discover how myTomorrows search could guide users through the maze of search to the desired information effectively and pleasantly.

Another problem posed is the lack of effective judgment or feedback from users. Implicit (Kelly et al., 2003) and explicit relevance feedback have been systematically studied and widely adopted, and recent research has gone creative in terms of interaction (Juan, et al., 2017). However, the lack of such data at myTomorrows interrupts the workflows of the internal medical team and the research team. Also, it is rather difficult to define the line where such data collection can meet AI optimization needs and is also acceptable by users. Hence, the question is how could relevance feedback with quality be collected in a data-sensitive and limited user scale context as myTomorrows?

Therefore, the goal of the research is to:

- 1) Understand current search flow and behavioral patterns of entire information-seeking cycles.
- 2) Explore interactions that may better guide users, engage more user judgments and measurably improve outcomes (e.g., efficiently leading users to relevant treatments).
- 3) Synthesize visions for the future of myTomorrows search and recommendation.

ASSIGNMENT **

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in "problem definition". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

This project expects two deliverables. One is a future concept design of myTomorrows search and recommendation system that enables better user guidance and engagement. The other is a supporting guideline or framework generalized from the research, which illustrates underlying principles and factors towards a clear and effective search experience.

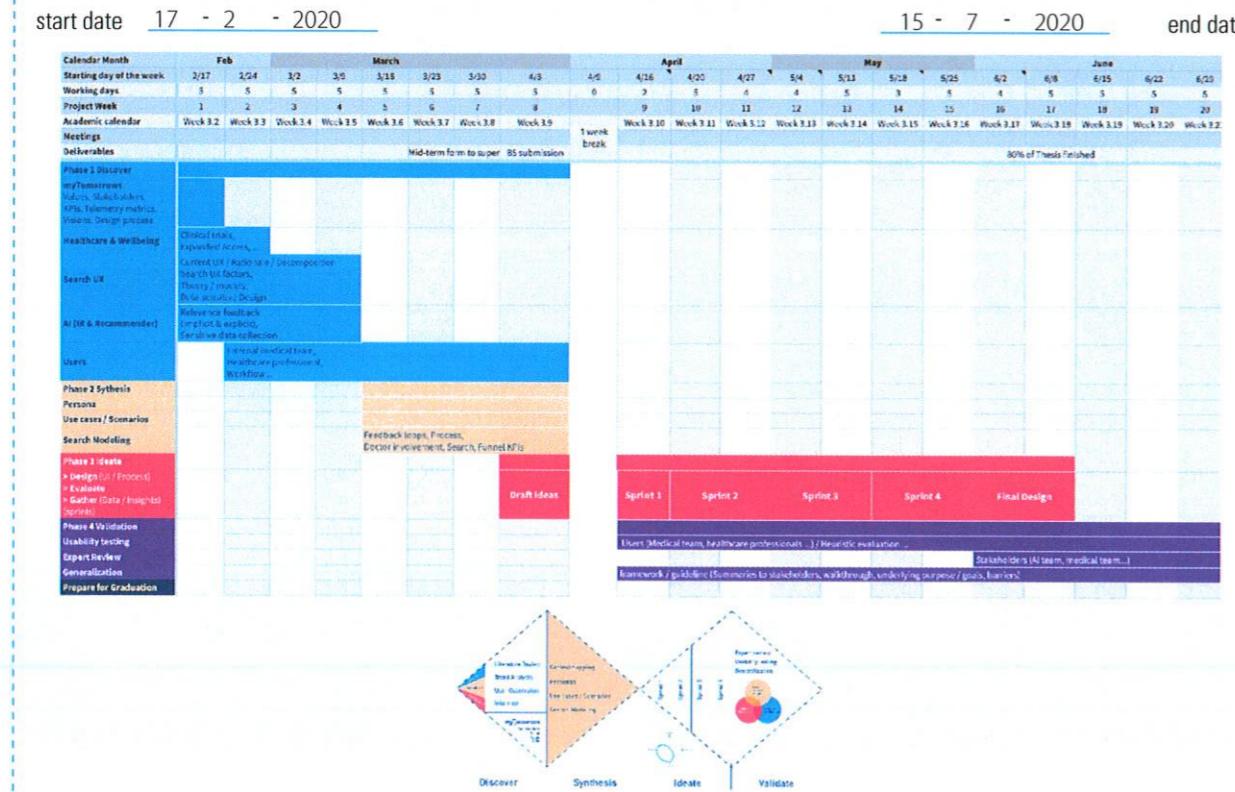
The goal is to create a clear and effective search experience. To achieve the intended effect, this project will employ Human-centered design methods to approach the problem. By synthesizing visions and incorporating theory and models to iteratively design, evaluate and gain feedback, this research aims at exploring interactions that may better guide and engage users, uncovering search factors affecting experience and measurably improving the experience of myTomorrows search.

The scope of research will cover contextual information, knowledge in academia and practice, and studies on users. More specifically in relevance feedback collection, investigations will be carried out into on-paper constraints, users' acceptance, and explicit AI optimization requirements. In order to reveal what can be gathered, how to gather through interactions without interrupting or freaking users, and find out the middle ground for the right collection of data.

In this project, research through design (Zimmerman, 2007) will be a highlighted method from which the insights and knowledge produced will lead to a validated design of search and recommendation system that is able to clearly and effectively guide users throughout search to desired information, yet gather relevance feedback that is trustworthy for users to share and reliable for AI optimization. Meanwhile, a supporting guideline or framework generalized from the research and design process will be structured as a reference for the future design of myTomorrows search experience.

PLANNING AND APPROACH **

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.



This project will run through four phases. In the first 8 weeks, it will be focused on collecting data, understanding the context, and building up search models for the next phase to take action upon. The following two phases will emphasize on design sprints to deepen the understanding of guiding users and relevance feedback collection. In the end, a concrete concept will be designed and validated, together with a generalized framework or guideline.

Discover: The research in this phase will start with the company research to understand its value, visions, stakeholders, and metrics. Followed by literature review and trend analysis in the intersection of Health & Wellbeing, AI (IR & Recommender) and Human-centered design. In addition, user research by observation and interview will be conducted to collect data on human factors.

Synthesis: With the raw data collected in the previous phase, I will use design tools such as Contextmapping, Personas, and Scenarios to structure and organize the data. Along with theory and models, the information and knowledge will form the foundation of design conceptualization.

Ideate: Ideating interactions to reach the ideal experience with Design Sprints (GV, 2016). In this phase, I expect four rounds of sprints to rapidly conceptualize, test and gain feedback. At the end of this phase, a concrete design idea will be defined, yet needs further evaluation and

Validate: Validate will focus on measuring the success and usefulness of the design with the involvement of users, experts, and company stakeholders. This phase will also generalize insights, factors, and search processes gained from previous research into a high-level framework or guideline for future reference.

MOTIVATION AND PERSONAL AMBITIONS

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge about a specific subject, broadening your competences or experimenting with a specific tool and/or methodology, Stick to no more than five ambitions.

As this project sits in the intersection of Health & Wellbeing, AI and Human-centered Design, it is a good starting point to deepen my knowledge in AI both theoretically and practically, meanwhile, systematically incorporate it into design practice. Search and recommendation system is available everywhere on the web nowadays but to the best of my knowledge, few aims at improving the access to Health & Wellbeing by now. I am thrilled to spare my expertise obtained from the Design discipline to inspire and contribute to this intersection aforementioned.

Personal ambitions:

- 1) Research and understand search and recommender AI systematically
Web search has been a part of my life ever since I started to interact with computers in my childhood. I search for answers to questions, inspirations, and expand my knowledge. Search looks so simple that it starts within a box of pixels and ends up with tons of information in milliseconds. How it works remains to be a black box to me. In this project, I would like to research, learn and uncover the mystery of searching from the technical, and human perspectives.
- 2) Harness design methodology and improve individual project management
Design tackles wicked problems in conjunction with other disciplines, and utilizing methodology to approach these problems serves as a fundamental skill for a design practitioner. Applying methodology with scalability to projects is what I desire to achieve, and develop a better way of working.
- 3) Enhance the skills of prototyping AI system and assessment
Prototyping is a quick way to test and examine ideas, while prototyping AI systems, traditional prototype tools such as sketch might not be supportive enough for quick prototyping and test. Therefore, I will try to improve my skills of web-based prototyping skills and ways of collecting data for assessment

FINAL COMMENTS

In case your project brief needs final comments, please add any information you think is relevant.

1.2.2 myTomorrows service funnel

At present, myTomorrows service funnel covers seven steps (Figure 4) and focuses on engaging both healthcare professionals (HCPs) and patients. Figure 72 integrates the existing funnel and maps out the user engagement journey.

Search report plays an essential role

Currently, the journey focuses more on the patients' end, and the fourth step (Inform) plays a vital role in user engagement. Based on requests, customized search reports (Figure 6, a PDF document that contains a list of investigational treatment options) will be produced by myTomorrows' medical team (internal HCPs) with its technologies. With the information provided in the search report, patients will be able to discuss treatment options with their treating HCPs on further actions to take.

Providing more than just information

However, whether enrolling a patient in a CT/EAP depends mainly on his/her treating HCPs' because of expertise and professional judgment required. HCPs' decision is inevitable, and it works as an essential part of the whole service funnel. Therefore, the future user engagement strategy will also emphasize the HCPs' end where myTomorrows digital product (a search engine for investigational treatment options) will play a key role, aiming to create an engagement loop and help more patients through this path.

	0
CT ClinicalTrials	0
EAP ExpandedAccessPrograms	1
RA RegulatoryArbitrage	0
OL Off Label Use	0

What are ClinicalTrials? What are ExpandedAccessPrograms?
What are ClinicalTrials? What are ExpandedAccessPrograms?

Figure 71. Screenshot of myTomorrows' search report (image source: myTomorrows)

Legend

- Patient engagement flow
- HCP engagement flow
- Service backend
- Conversion

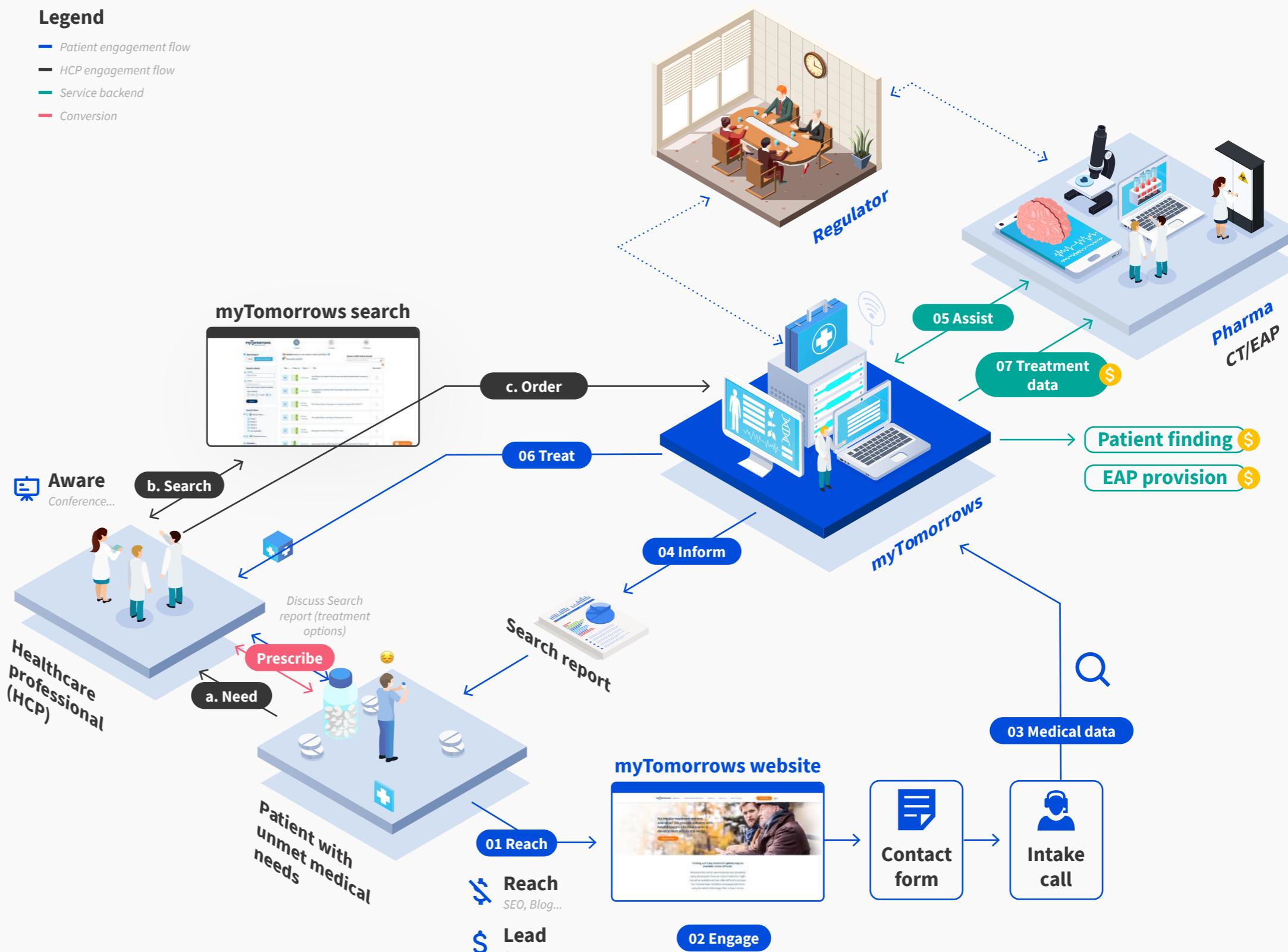


Figure 72. HCP-oriented service system map (image drawn by author)



Improving Search Relevance Feedback through Human Centered Design

DfI Master Thesis