How to Research

Luca Maria Aiello
Head of Study Program for KDS
IT UNIVERSITY OF CPH

Today's objective

Provide you with a (motivated) checklist of best practices to use during your Research Projects

Point A:
No idea what to do

Many of you are probably somewhere here

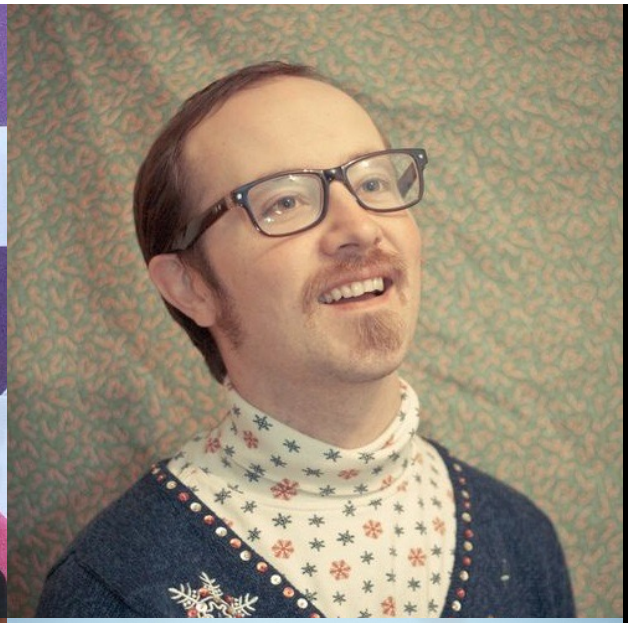
Point B:
A plan of action to follow

Who gives you the authority?!?!

- In the Research Business since 2009
- Industrial & academic setting
- 93 peer-reviewed publications
- A copious basket of failed papers
- Avg. 1 publication every ~1.8 months

It’s not about the “medals”, it’s about the long-time experience in multiple settings

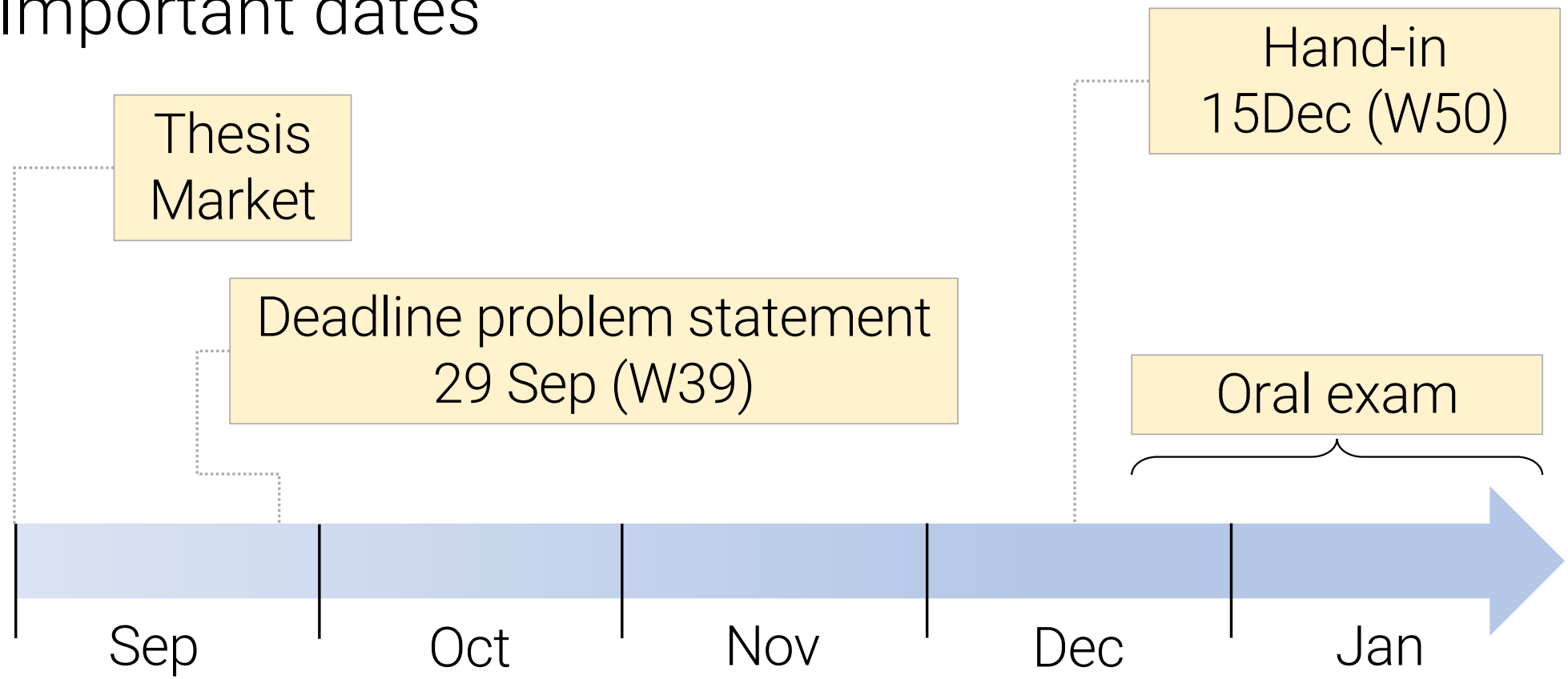
Beware of Survivor Bias

FOR BTS V


Disclaimers

- No guaranteed nor exhaustive recipe for Research
  - But… some practices generalize quite well and are widely adopted
- Most of you won’t be Researchers by profession
  - But… acquiring a Researcher mindset will help you go “above & beyond”
- Research is a collaborative effort
  - But… making it work is an individual responsibility

Important dates

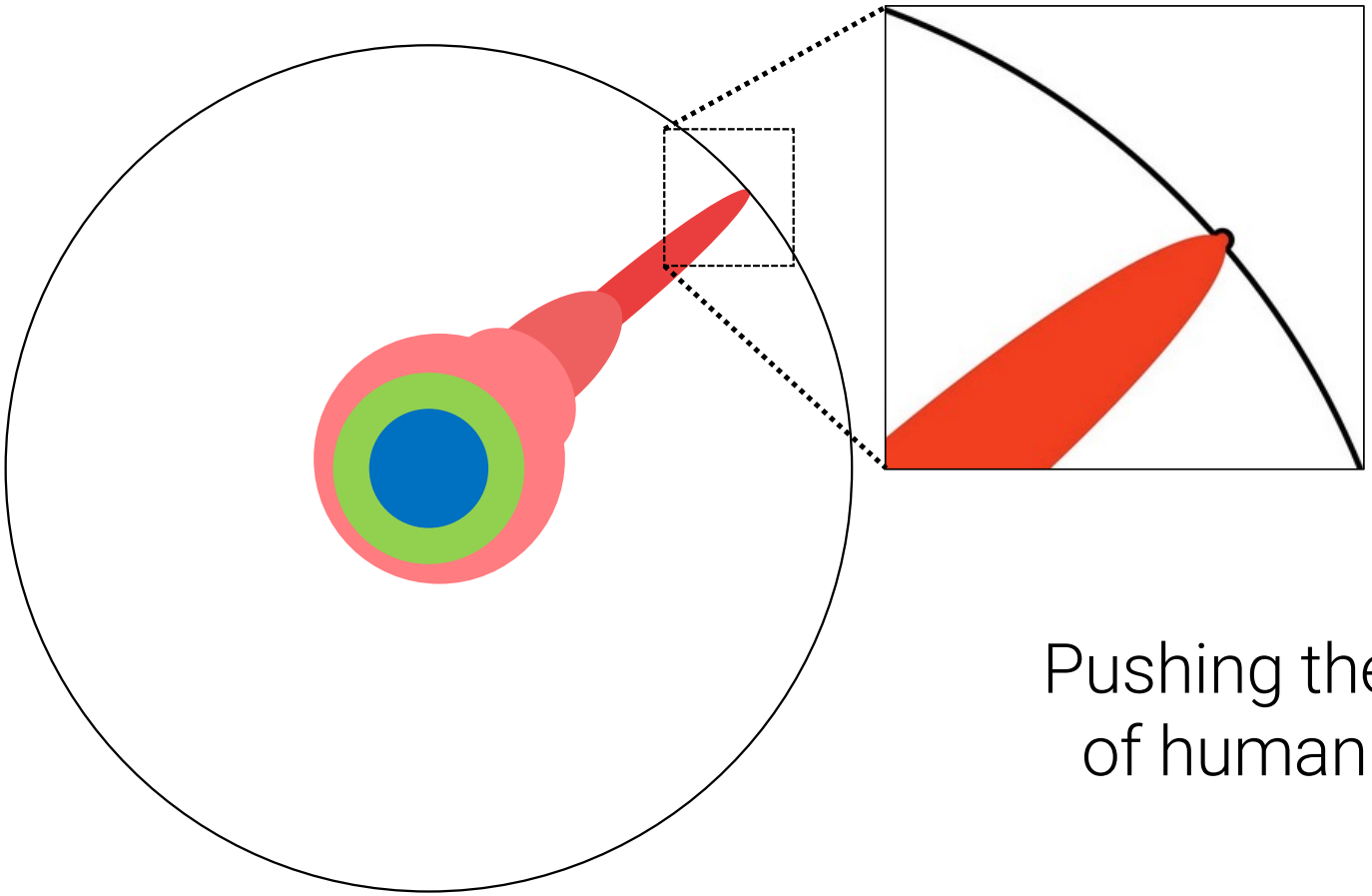


Process steps:
https://itustudent.itu.dk/Study-Administration/Project-Work/Process-for-writing-a-project-thesis

Program

- What is Research?
- Find the Research Question
- Assemble a Group
- Turning the Question into a Research Project
- Literture review
- Break
- The Mentor & the Agreement
- Time Management
- Writing
- Dealing with failure
- Open discussion

What is Research?

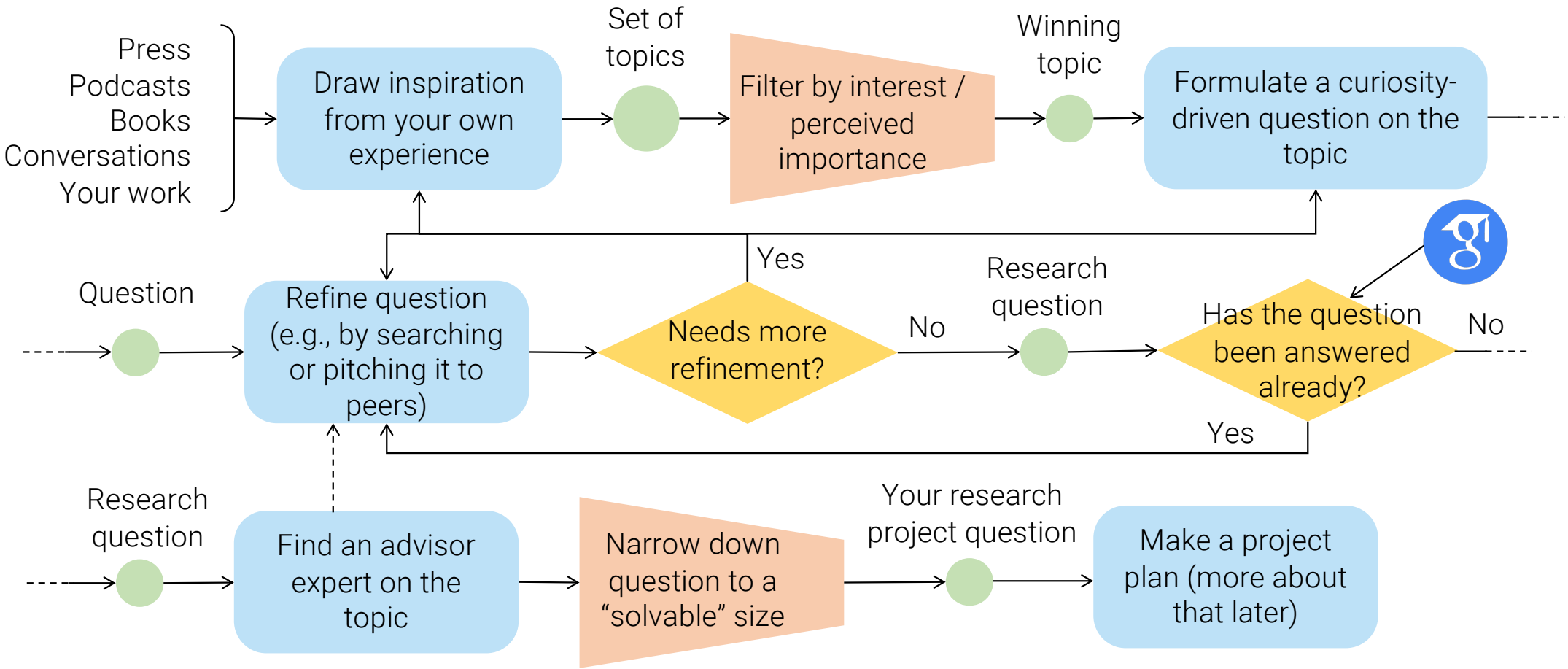


Pushing the boundaries
of human knowledge

Matt Might, The Illustrated Guide to Ph.D : http://matt.might.net/articles/phd-school-in-pictures/

The Research
Question

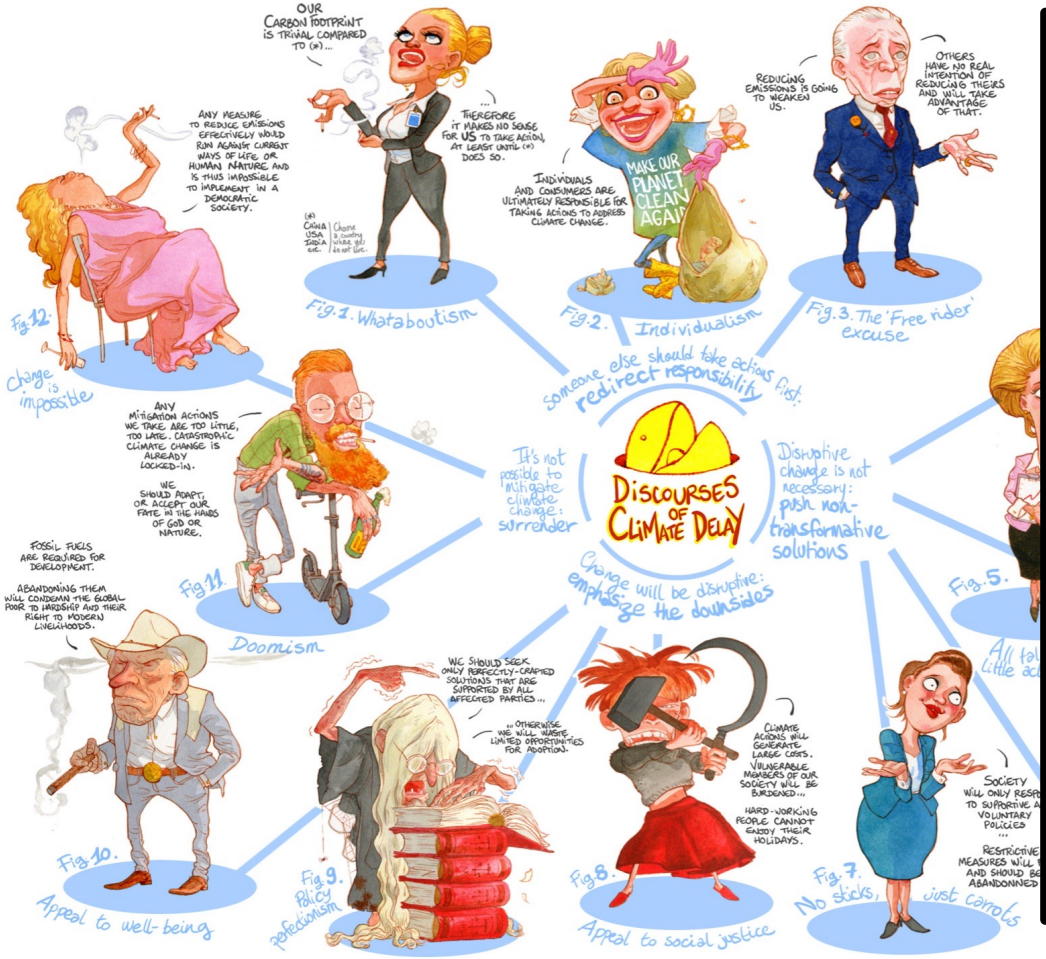
How to find a Research Question (ideal scenario)

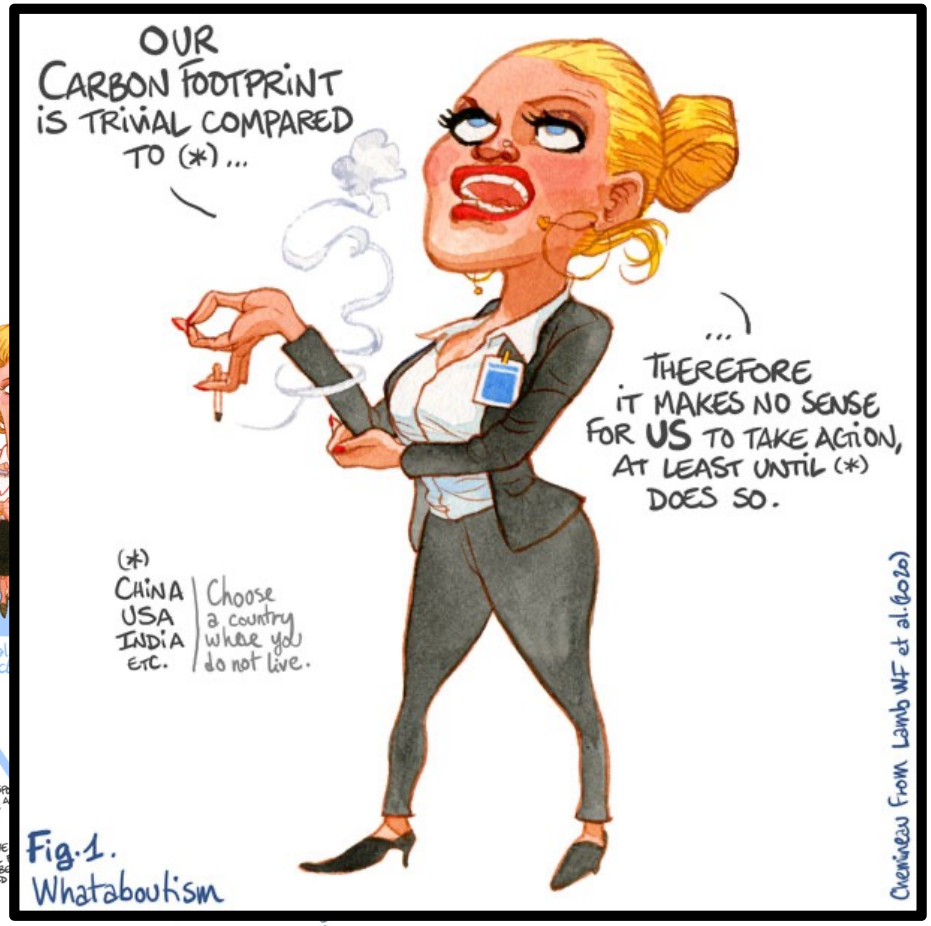


Example (trivialized)

- The news is all about Ukraine, climate change, and Musk Inspiration from news
- I am worried about climate change as I fear for the hardships my kids may need to endure one day Filter by passion on topic
- I wonder why people around me still don't seem to care about doing anything in their daily lives to tackle global warming Curiosity-driven question
- "Why people delay/avoid climate action?" Refine question
- Result #1 on Google Scholar is a paper that explain 12 common reasons for climate inaction Check for answers
- Do we know how common these 12 reasons are? Refine question
- Google Scholar shows only a bunch of small-scale studies Check for answers
- "How common are the 12 arguments for climate inaction in public discourse?" Research question

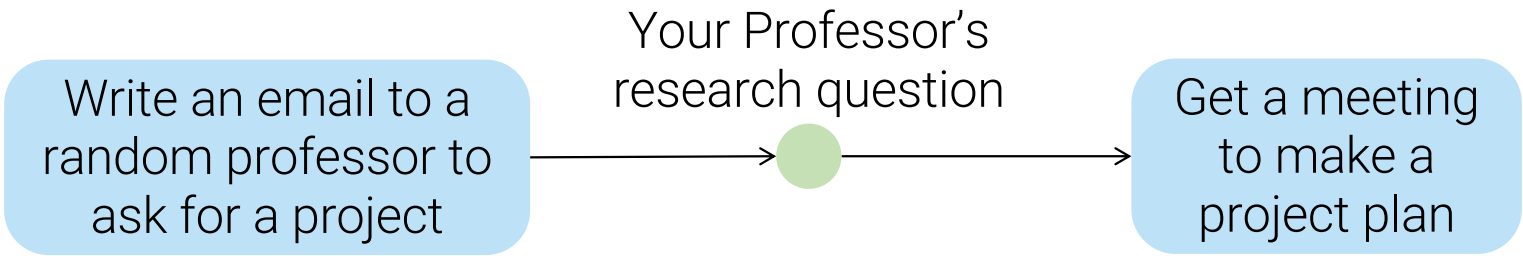
Discourses of climate delay





Lamb et al. "Discourses of Climate Delay", 2020

How to find a Research Question (quick path)



Note: it's still totally OK to ask professors for ideas early on in the process, and get inspired by the ideas they pitch to you (e.g., thesis market).

Keep in mind

- Don’t need to know previous research to formulate a question
- Get inspired by your experience (in class, at work, in life)
- Discuss your ideas with others

Assemble a group

Assemble a group

- We advise to do your project in groups of 2 or 3
  - More interesting/ambitious project
  - More sustainable workload
  - Diversity of expertise and attitude
  - More fun
  - More coordination cost that you will have to handle
- Remember, grades are individual
- If your group has students from more than one programme in it, you must inform the relevant HoSPs (Marco, Luca, Patrick).

Turning your Question into a Research Project

The Research Proposal

Expand your question into three parts:

- **Situation (Context)**
  - Describe the problem and motivate its importance
- **Complication (Gap)**
  - Explain why the problem hasn't been fully solved yet
- **Proposal (Innovation)**
  - Propose a new solution that solves (part of) the problem

To draft these, knowledge of previous research in the topic is needed

The Research Proposal

"How common are the 12 arguments for climate delay in public discourse"

Situation

Need support

Humankind needs to adopt widespread lifestyle changes to avert catastrophic consequences of global warming. To nudge people into new habits, it is key to know the reasons that keep people away from behavioural change.

Need support

Complication

Result of previous literature search

Previous work has identified 12 main arguments that may be used to avoid or delay climate action. However, the relative importance of these arguments in the public discourse is unknown, because it is hard to quantitatively measure their expressions from natural language.

Existing related work?

Proposal

We propose to train a Natural Language Processing tool to identify automatically the 12 arguments of delay from conversational text, and use it to measure their relative frequency of occurrence in Twitter conversations

The Research Proposal

Situation

Machine vision systems are vulnerable to adversarial attacks. Exploring all the possible vulnerability scenarios is key to assess the resilience of these systems

Complication

Previous work have assumed a threat model in which the adversary can digitally modify the image. This does not consider the scenario of adversarial examples introduced in the real world, in the scene captured by the machine vision system

Proposal

We propose to print physical “adversarial patches” and measure their effectiveness in perturbing classification accuracy when introduced in a scene, under different conditions

Inspired by: Kurain et al. “Adversarial examples in the physical world”, 2017

The Research Objective

- Establish a precise, quantifiable objective for your research

"How common are the 12 arguments for climate inaction in public discourse?" → *"For each of the 12 arguments, measure the fraction of social media posts that mention them during the last year"*

- The objective will guide you in the definition of your methodological framework

The Methodological Framework

Define in this direction

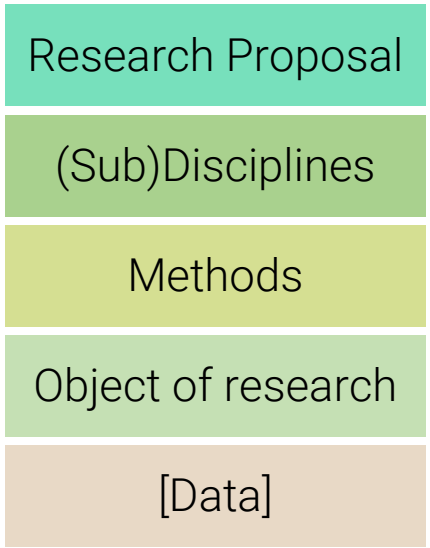
Research Proposal	
(Sub)Disciplines	Sets of tools, pratices, formalisms, literature, etc. The same Question can be approached from multiple angles and different disciplines
Methods	Usually, the Discipline indicates possible methods to answer Question. New methods can be created if needed/relevant
Object of research	The (abstract) object the methods will be applied on
[Data]	Data that instantiates the object of study (not always relevant, e.g., theoretical research)

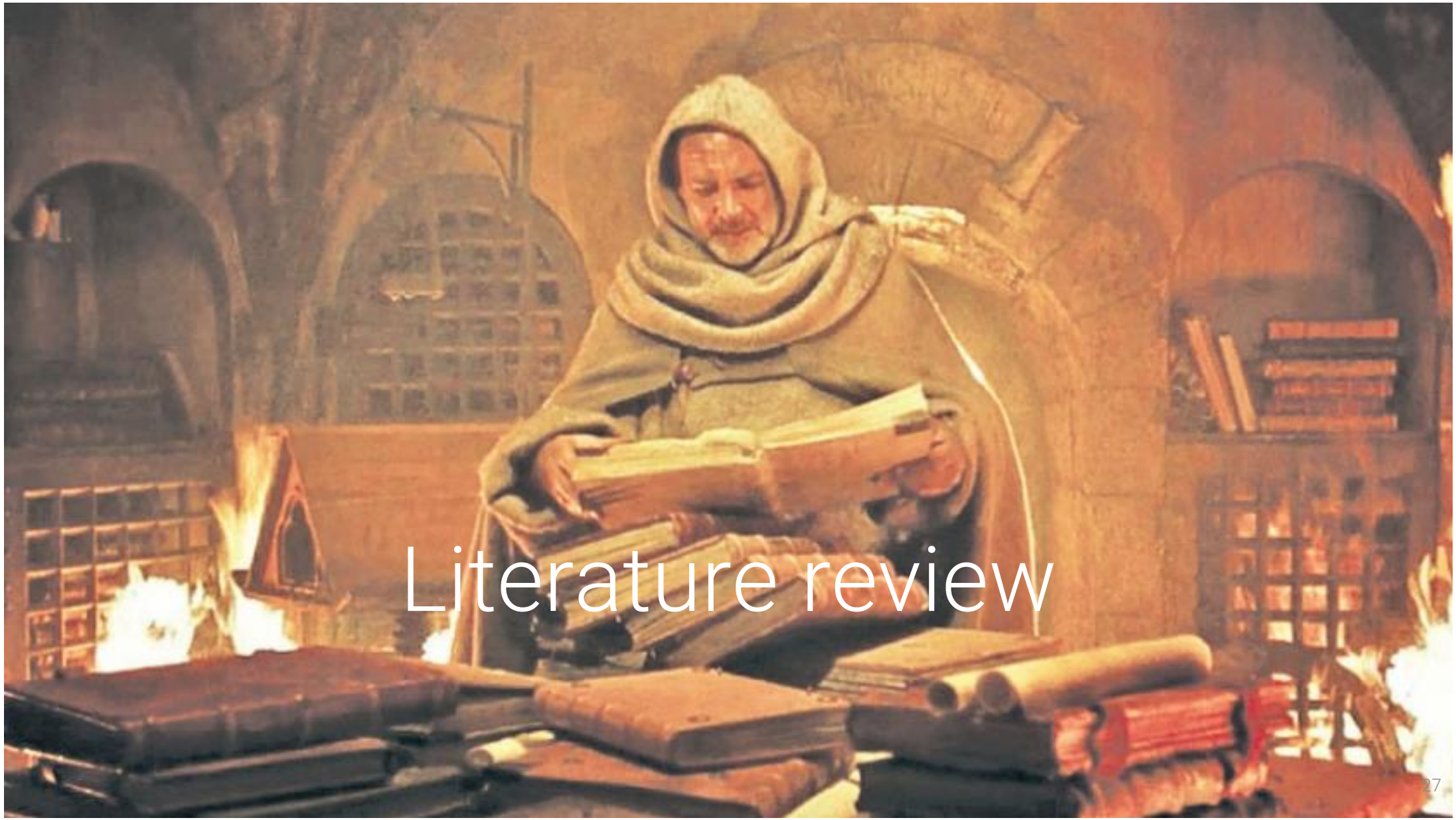
The Methodological Framework

Research Proposal	*"Measure frequency of 12 arguments for climate delay in public discourse"*
(Sub)Disciplines	~~Ethnography~~ Computational Social Science ✓ Natural Language Processing
Methods	Transformer models for classification Fine-tuning on novel training data
Object of research	Online conversations ✓ ~~News~~ media
[Data]	Mastodon ✓ ~~Reddit~~

Keep in mind

- Try to resist the natural temptation to start a research project from methods or data
  - “LLMs are cool, I want to do something with them!”
  - “I want to mine some TikTok data!”
- It’s ok to be inspired by cool methods or datasets, but try to formulate a Research Question as soon as possible
- Establish a quantitative objective, craft your methodology around it





Literature review

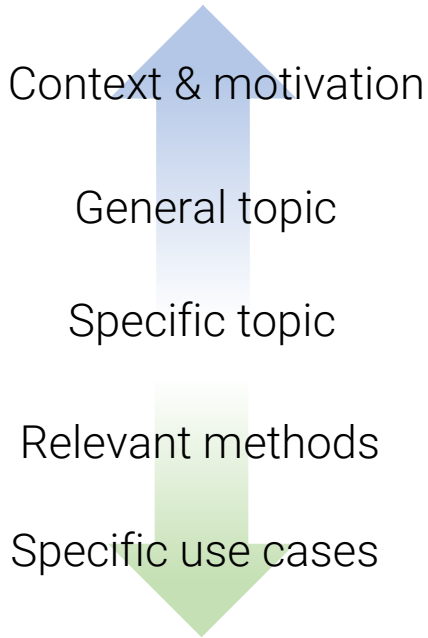
Literature review criteria

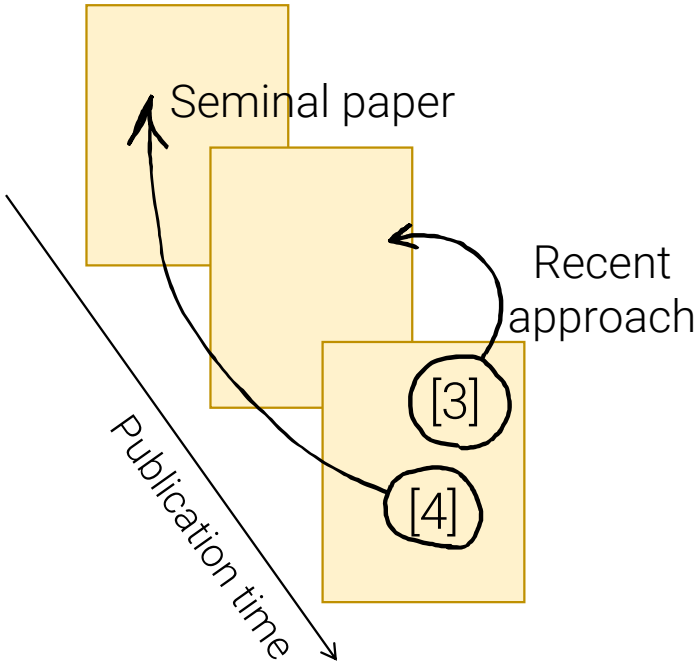
- Relevance
  - Submit multiple queries to specialized search engines (e.g., Google Scholar)
  - Filter by relevant venue
- Impact (quality)
  - Highly-cited papers
  - Papers published in top-tier venues
  - Papers published by top authors in the field
  - Papers recommended by experts
- Recency
  - To make sure your proposal is novel, search up to most recent publications

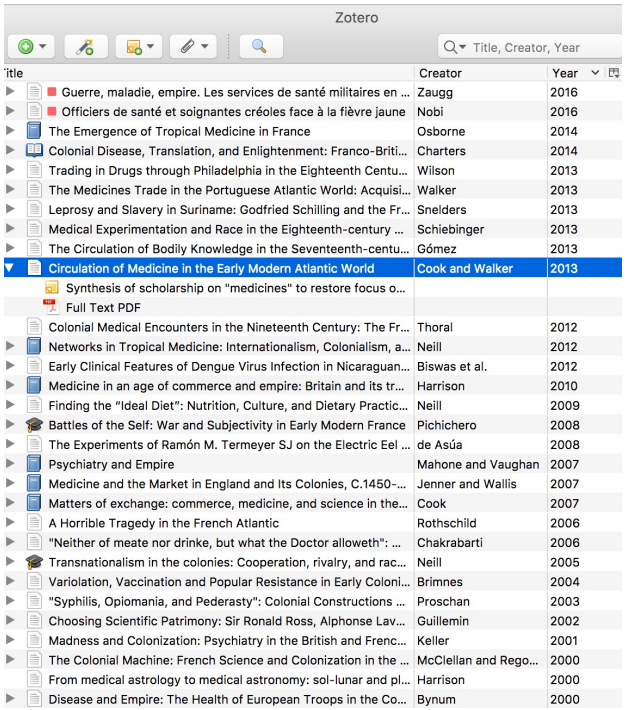
How do I know?

With practice, you'll know.
Until then, ask your mentors

Literature search tips







Title	Creator	Year
Guerre, maladie, empire. Les services de santé militaires en ...	Zaugg	2016
Officiers de santé et soignantes créoles face à la fièvre jaune	Nobi	2016
The Emergence of Tropical Medicine in France	Osborne	2014
Colonial Disease, Translation, and Enlightenment: Franco-Briti...	Charters	2014
Trading in Drugs through Philadelphia in the Eighteenth Centu...	Wilson	2013
The Medicines Trade in the Portuguese Atlantic World: Acquisi...	Walker	2013
Leprosy and Slavery in Suriname: Godfried Schilling and the Fr...	Snelders	2013
Medical Experimentation and Race in the Eighteenth-century ...	Schiebinger	2013
The Circulation of Bodily Knowledge in the Seventeenth-centu...	Gómez	2013
Circulation of Medicine in the Early Modern Atlantic World	Cook and Walker	2013
Synthesis of scholarship on "medicines" to restore focus o...		
Full Text PDF		
Colonial Medical Encounters in the Nineteenth Century: The Fr...	Thoral	2012
Networks in Tropical Medicine: Internationalism, Colonialism, a...	Neill	2012
Early Clinical Features of Dengue Virus Infection in Nicaraguan...	Biswas et al.	2012
Medicine in an age of commerce and empire: Britain and its tr...	Harrison	2010
Finding the "Ideal Diet": Nutrition, Culture, and Dietary Practic...	Neill	2009
Battles of the Self: War and Subjectivity in Early Modern France	Pichichero	2008
The Experiments of Ramón M. Termeyer SJ on the Electric Eel ...	de Asúa	2008
Psychiatry and Empire	Mahone and Vaughan	2007
Medicine and the Market in England and Its Colonies, C.1450-...	Jenner and Wallis	2007
Matters of exchange: commerce, medicine, and science in the...	Cook	2007
A Horrible Tragedy in the French Atlantic	Rothschild	2006
"Neither of meate nor drinke, but what the Doctor alloweth": ...	Chakrabarti	2006
Transnationalism in the colonies: Cooperation, rivalry, and rac...	Neill	2005
Variolation, Vaccination and Popular Resistance in Early Coloni...	Brimnes	2004
"Syphilis, Opiomania, and Pederasty": Colonial Constructions ...	Proschan	2003
Choosing Scientific Patrimony: Sir Ronald Ross, Alphonse Lav...	Guillemin	2002
Madness and Colonization: Psychiatry in the British and Frenc...	Keller	2001
The Colonial Machine: French Science and Colonization in the ...	McClellan and Rego...	2000
From medical astrology to medical astronomy: sol-lunar and pl...	Harrison	2000
Disease and Empire: The Health of European Troops in the Co...	Bynum	2000

Start relatively narrow on target topic. Broaden for context and motivation, narrow for existing solutions

Snowball search backward in time

Keep track and annotate with tags for grouping and later retrieval (e.g., Zotero)

Literature review – how to find if paper is relevant?

- Don’t need to read them fully

- Read title and abstract
- Skim introduction, focusing on last part listing contributions
- Skim conclusions

- Discard at any point if clearly irrelevant
- Take notes for indexing/clustering

“I found 100 relevant papers, there’s no way I can read all of them”

Keep in mind

- Literature search can go on forever, set a stopping criterion
  - E.g., the last 3 Google Scholar result pages on each query have 0 hits

- You will likely find lots of work on your topic of interest (unless you are exploring a very frontier topic)

- Don't feel overwhelmed by what's already done
- Anything that "pushes the boundary" is research, remember?
- Incremental questions are OK! Most research is incremental…

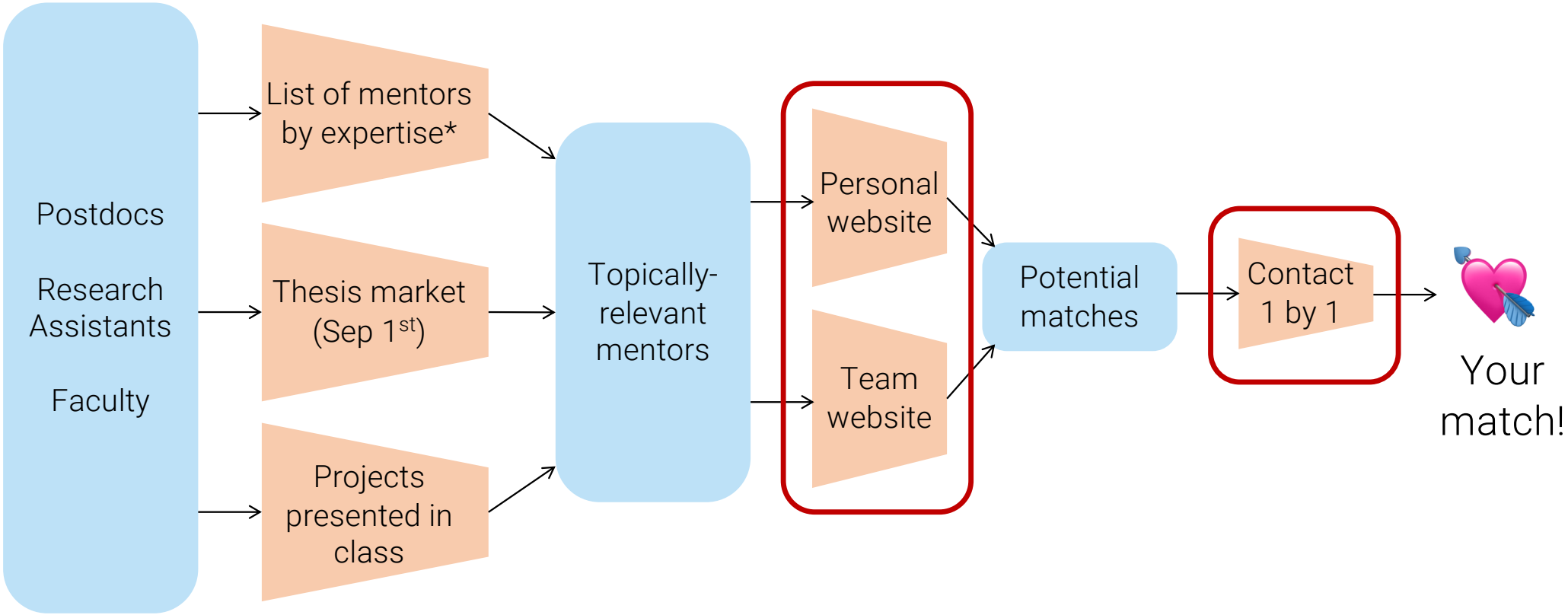
Good free Coursera course on Literature review:
https://www.coursera.org/learn/academicinfoseek

Short break

- What is Research?
- Find the Research Question
- Assemble a Group
- Turning Question into Project
- Literture review
- Break
- The Mentor
- Time Management
- Writing
- Dealing with failure
- Open discussion

The Mentor

The right match (at ITU)



* https://itustudent.itu.dk/Study-Administration/Project-Work/Finding-a-supervisor

Personal & team website (know your potential mentors)

Students

Welcome to the NERDS side! 🤓
We are always interested in working with motivated and dedicated students and have plenty of interesting opportunities for Master and Bachelor theses. If you are interested in doing a project with us please follow the 4 steps below.

1) Before you contact us: Are we a match? 💘

Our general expertise best fits the Data Science study programs, but we are happy to supervise students from all study programs if the topic makes sense. Make sure to explore the full list of computer science advisors at ITU before approaching us – someone else might be a better fit for you: https://wiki.itu.dk/computerscience/index.php/Supervisors_for_projects_and_theses

We strongly prefer to supervise groups of 2 or 3 students.

🚨🚨🚨 **Before** you approach us, make sure the following know-how exists in your group:

- 10 must-know statistical concepts
- Data visualization basics
- Python or R programming
- LaTeX
- If you plan to work with networks: Networks basics (pdf)

Requirements listed

HOME | TEACHING | PUBLICATIONS | CODE | DATASETS | CV | CONTACT | REPRO

Project Ideas

Below are some ideas for Master's theses, bachelor projects and other projects for students at the IT University of Copenhagen. Contact me (mcos@itu.dk) to get started.

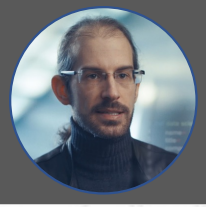
- **Efficient estimation of a network's polarization**: I have developed a measure to estimate how polarized the political discourse is on Facebook and Twitter. However, the measure can only be calculated efficiently for small networks. In this project, you will explore new algorithms to improve the efficiency of the calculation.
- **Affective polarization on social media**: having different opinions is not an issue, but it is when it causes people to talk to other people in aggressive way. This is affective polarization: the perception of disagreement as an enemy action. Can we estimate the level of affective polarization on a social network? How does it relate to opinion polarization?
- **Build Multilayer Networks from Graphic Novels**: making a network out of a work of fiction is helpful to study literary theory and gender norms in fiction. Most works so far build networks with a single type of interaction, but multilayer and signed network could potentially be more useful. A good potential material could be Kentaro Miura's Berserk, since it contains different arcs and characters changing allegiances.
- **Human mobility and development**: Using a new dataset estimating the development level of places in developing countries, we can estimate the link between wealth and the place

More project listed

Contact your Mentor. Email should contain:

Dear [Charles | Dear Prof. Xavier],	Greeting
My name is Jimmy Morales, a final-year student at KCS.	Self-introduction
I am searching for a mentor for my Master Thesis. During my Master I have developed an interest in the topics of security and cryptography, and I would be very interested in investigating security vulnerabilities of NFT smart contracts.	Research objective
Your recent work on NFTs [link] resonates with the research idea I'd like to develop.	Reason of perceived match
In case this is of interest, would you have time for a quick chat to discuss options?	Proposed action

Bad example (trivialized)



Professor Mc Professorface
Last seen today at 9.15 am

Hej wassup dawg, I need a real solid. Got any spare reseach project that u can hand out??? Need one ASAP to get my missing credits!!
3.24 am

Well… pass by my office next week and we'll see what we can do
8.31 am

The Project Agreement (Preliminary Project Statement)

- Students **must** submit a Preliminary Problem Statement for their supervisor's approval

- The statement specifies the scope of the project, a description of the proposed subject area, topic, hypothesis etc.

- The Project Statement is submitted on LearnIT (a Project page will be available on Week 38)

https://itustudent.itu.dk/Study-Administration/Project-Work/The-project-page
How to write a project statement: http://www.itu.dk/people/wasowski/papers/project_guide.pdf

The Supervision Agreement

- Optional, supplement to project statement
- **Highly** recommended
- Purpose: align expectations
  - What your mentor can do for you
  - Define objectives
  - Meetings, feedback, deadlines…
  - **Agree on use of generative AI tools**

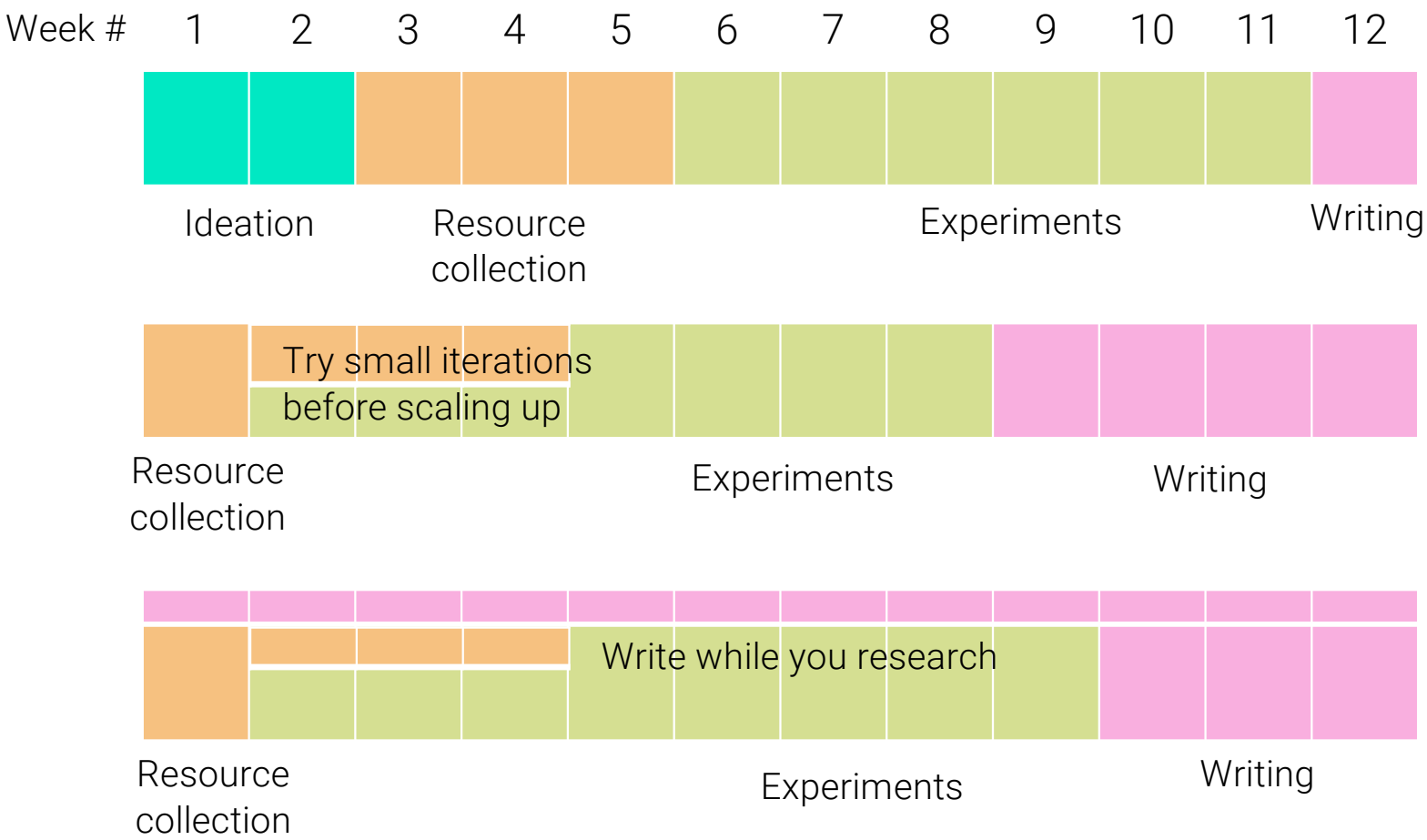
https://itustudent.itu.dk/Study-Administration/Project-Work/Your-supervisor-and-supervision
https://intranet.itu.dk/Quality-in-Education/Supervision-evaluation/Supervision-agreement
Collaboration policies & AI: https://github.itu.dk/thore/itu-cs-collaboration

Keep in mind

- Evaluate match in terms of the similarity of the mentor's expertise/focus with your research question
- Junior researchers (e.g., postdocs) can be great mentors
  - Hands-on, fresh perspective on their core topic
- More than one match? Excellent starting point for co-supervision
- Agree on expectations early on with explicit agreements

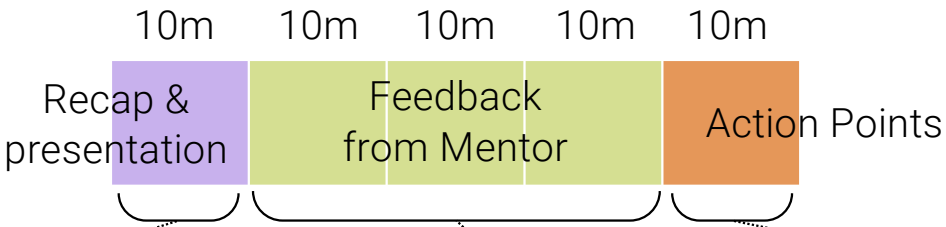
Time management

Project timeline examples



Examples given with a Data Science project in mind. Proportions may vary in other fields.

Managing supervision time effectively

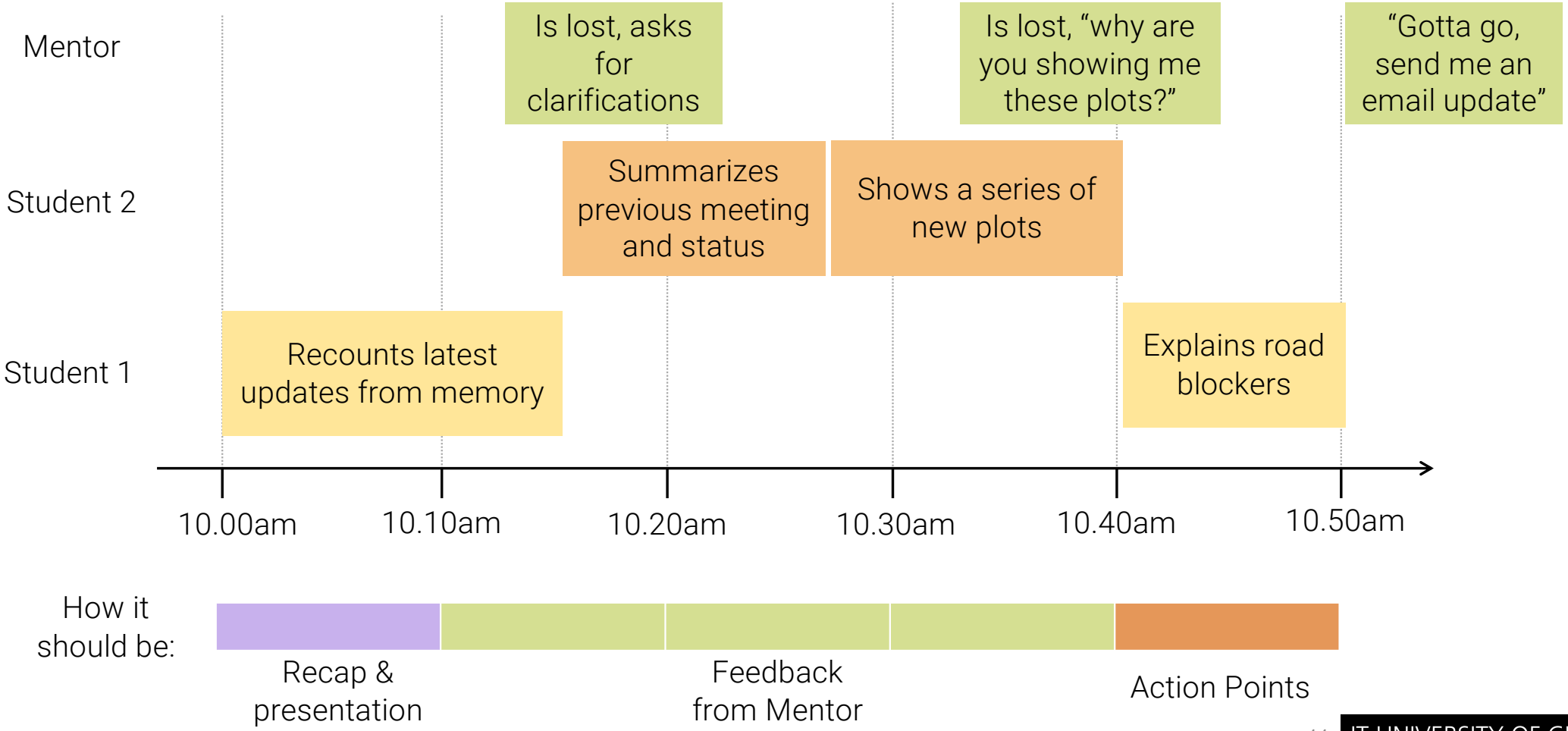


- Assume Mentor is memoryless
- Recap information relevant to decisions to be taken
- Present previous action points and how they were (not) addressed
- Use a structured presentation medium (ugly ppt is good enough) instead of improvising

- Be explicit about request to Mentor to avoid aimless feedback
- Take extensive written notes (anything that is not written will be forgotten)
- Take **all** feedback very seriously

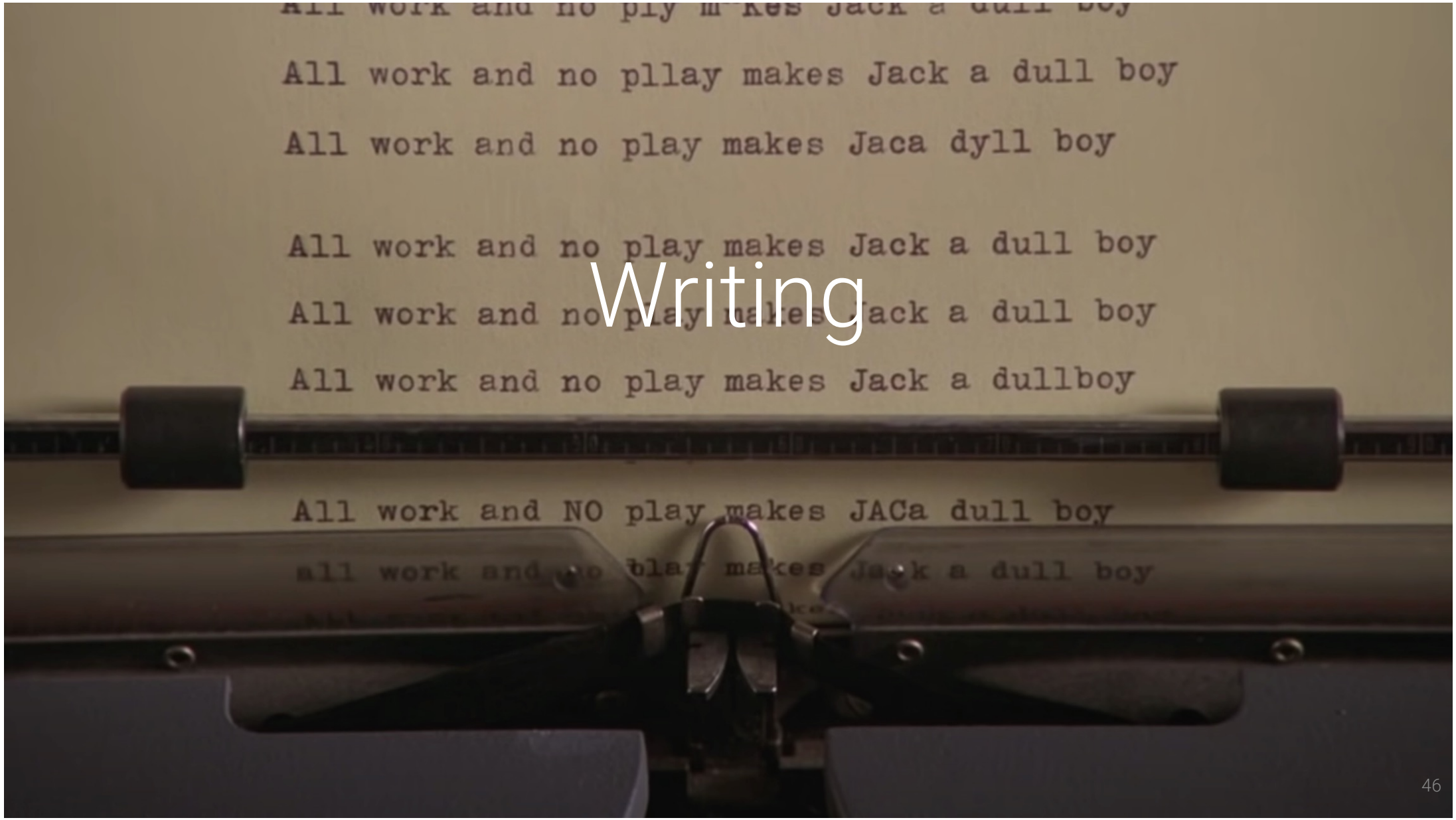
- Agree on action points and note them down
- Take time to make sure there is a common understanding of what action points are

Mentoring meeting going South (trivialized example)



Keep in mind

- Work in parallel (e.g., writing + experiments)
- Work iteratively (e.g., don’t wait for all data to download before experimenting)
- Keep detailed track of progress and no one will get lost
- Don’t underestimate the importance of writing, and the time it takes
- Contrary to expectations, YOU should manage supervision time to make the best out of the interaction with your Mentor
- Mentoring time is extremely precious: allocate time for meeting preparation to get the best out of the meeting

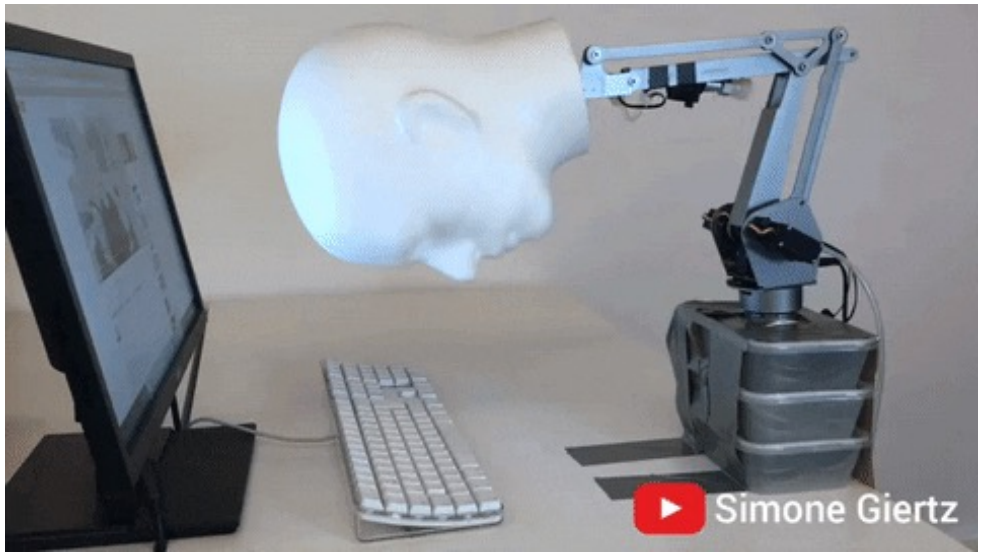

All work and no pllay makes Jack a dull boy
All work and no play makes Jaca dyll boy
All work and no play makes Jack a dull boy
All work and no play makes Jack a dull boy
All work and no play makes Jack a dullboy
All work and NO play makes JACa dull boy

Writing

GPT will just do the writing for me, right?

- Yes and no…
- Use it to the extent allowed it by the Supervision Agreement
- Acknowledge it fully and in detail

- You should always know how to write a good thesis to evaluate if your favorite AI did a good job



Possible paper structure* (10 pages, double column)

Ideally, early drafts of abstract/intro should be written early on, to guide the research process

Order	Section	Length	Readership
7	Title	1 line	10000
6	Abstract	1 paragraph	1000
3	Introduction	1-2 pages	100
1	Methods	4-6 pages	2
2	Results		5
4	Related Work	1-2 pages	25
5	Conclusions	1 page	50
	References	1 page	

Most readers won't go past the abstract

With practice you will learn to swap them

Yes: at the end of the paper, not to disrupt the flow between the intro and the "meat" of your contribution

*Format may vary depending on field. Discuss it with your supervisor

Abstract

- Situation
- Complication
- Proposal
- Contribution

This wasn't present in your research proposal – you did not know the results back then!

Marker often used to start the proposal

Writing concurrent programs is notoriously difficult, and is of increasing practical importance. A particular source of concern is that even correctly-implemented concurrency abstractions cannot be composed together to form larger abstractions. In this paper we present a new concurrency model, based on transactional memory, that offers far richer composition. All the usual benefits of transactional memory are present (e.g., freedom from deadlock), but in addition we describe new modular forms of blocking and choice that have been inaccessible in earlier work.

No references, unless absolutely necessary

Harris et al. "Composable memory transactions", 2005

Introduction

- Situation
  - Expand with motivation (+references)
- Complication
  - Expand with short description of problem and failed attempts (+references)
- Proposal
  - Expand with simple key ideas: re-usable insights useful to the reader
- Contribution
  - Report key findings with hard numbers. Bullet-point lists of contributions is common

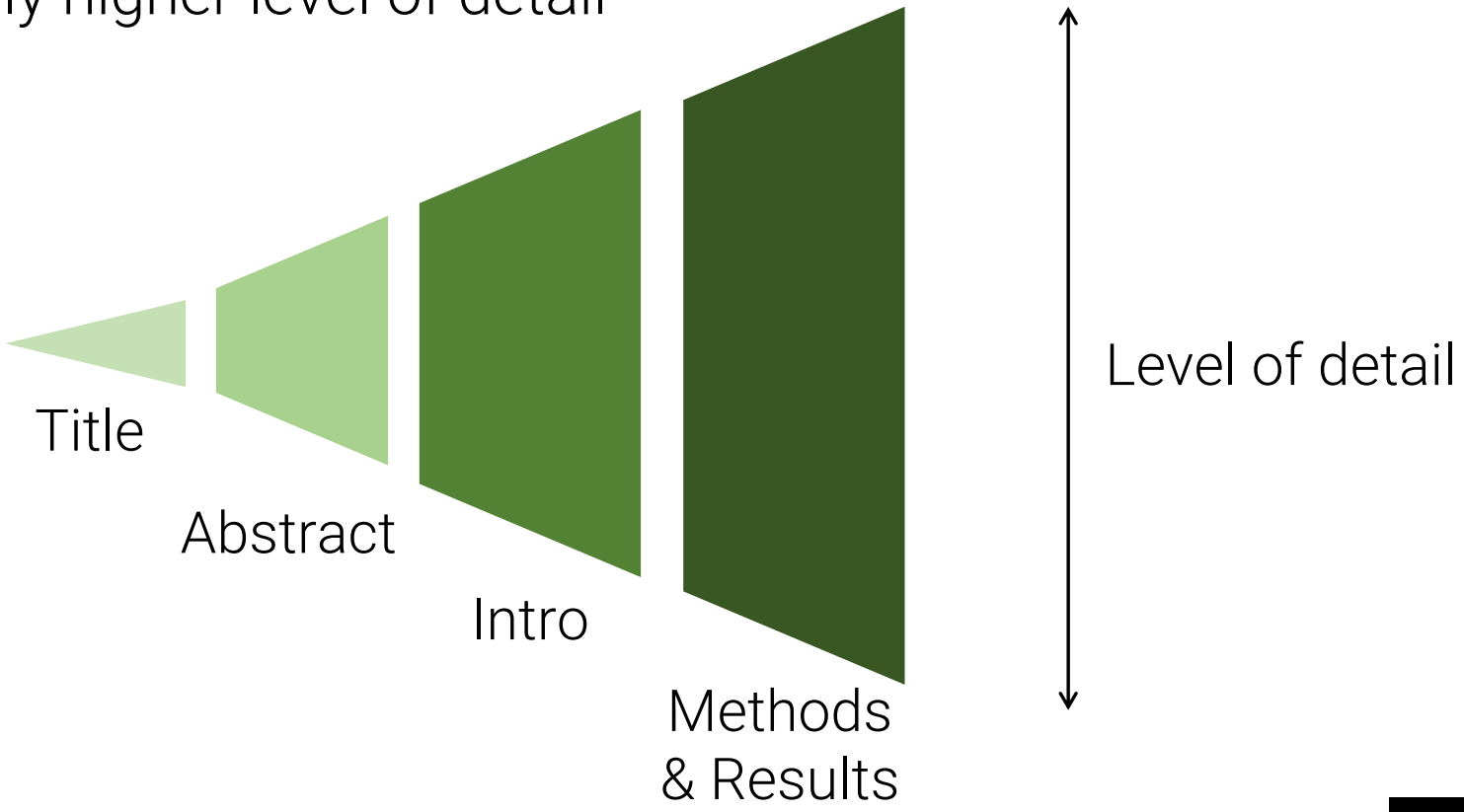
If the target community is very focused, the situation could be omitted (pro-writers)

Ideally, reading the introduction should be a convenient alternative to reading the whole paper

Derek Dreyer "How To Write Papers So People Can Read Them"
https://www.youtube.com/watch?v=aZp4vZccI6Q

Top-down exposition, layered presentation

Re-iterate the same idea with
increasingly higher level of detail



Related work

- Related work should be related to the situation or complication
- Provide real comparisons, not a laundry list
- Give credit, don't discredit

- How to select appropriate reference among many?
  - One point one reference
  - The reference that is more specifically related to your claim wins
  - Weigh prestige in

Simon Peyton Jones "How To Write a Great Research Paper"
https://www.youtube.com/watch?v=VK51E3gHENc

High-level checklist: Flow

- Each sentence/paragraph needs to be related to adjacent ones
- Old to new: begin sentences with old information, end them with new information

Security proofs of cryptographic protocols are crucial for the security of everyday electronic communication. However, these proofs tend to be complex and difficult to get right. The game-playing technique, originally proposed by Jones et al., follows a code-based approach where the security properties are formulated in terms of probabilistic program, called games. This is a general design principles for cryptographic proofs to ease their management.

Derek Dreyer "How To Write Papers So People Can Read Them"
https://www.youtube.com/watch?v=aZp4vZccI6Q

High-level checklist: Flow

- Each sentence/paragraph needs to be related to adjacent ones
- Old to new: begin sentences with old information, end them with new information

Security proofs of cryptographic protocols are crucial for the security of everyday electronic communication. However, these proofs tend to be complex and difficult to get right. To make it easier to manage such proofs, Jones et al. have proposed a new design principle called game-playing technique. This technique follows a code-based approach where the security properties are formulated in terms of probabilistic program, called games.

Derek Dreyer "How To Write Papers So People Can Read Them"
https://www.youtube.com/watch?v=aZp4vZccI6Q

High-level checklist: Coherence

- Each paragraph needs to make exactly one point
- The point made at the beginning of the paragraph
- The point is then developed in the rest of the paragraph

Lions and tigers are some of the most dramatic and awe-inspiring species of cats. Most of these large cats, however, are currently facing extinction. A smaller cat that has been more evolutionarily successful is the house cat. Although house cats are currently the most popular pet in the world, they are in many ways anti-social. It would therefore be interesting to study whether house cats can be training to be more sociable

Derek Dreyer "How To Write Papers So People Can Read Them"
https://www.youtube.com/watch?v=aZp4vZccI6Q

High-level checklist: Coherence

- Each paragraph needs to make exactly one point
- The point made at the beginning of the paragraph
- The point is then developed in the rest of the paragraph

There appears to be a negative correlation between the charisma of a species and its ability to survive. Lions and tigers, for instance, are among the most majestic creatures in the animal kingdom yet they are currently facing extinction. In contrast the house cat is known for being evolutionarily successful, even if it's not as magnificent as big cats.

Derek Dreyer "How To Write Papers So People Can Read Them"
https://www.youtube.com/watch?v=aZp4vZccI6Q

High-level checklist: pro tips

- “Name your baby”
  - Give names to things you use frequently
  - At the same time, try to avoid overuse of acronyms
- “Just in time”
  - Give information when is needed, not before
  - Check: is this concept useful in this or the next paragraph?

Derek Dreyer “How To Write Papers So People Can Read Them”
https://www.youtube.com/watch?v=aZp4vZccI6Q

Low-level checklist

Mircea Lungu "Writing Guidelines"
https://github.com/mircealungu/student-projects/blob/master/writing_guidelines/README.md

- After you have written a section, parse each rule in this list and check if your text complies with them

- "Every Figure Should be referenced in the text"
- "Use active voice"

~~The choice became Cloud Firestore as it …~~

We chose Cloud Firestore because…

- "Write facts, not opinions"

Another model by Authorson ~~is more promising. It~~ focuses on three areas...

Get feedback (and listen to it!)

- From your mentor
- From your peers
- From the community
  - arXiv + tweeting does wonders

Keep in mind

- Follow pre-determined paper structure
- Start writing early
- Just do it – write “bad” text, iterate to improve
- Scan each paragraph to check compliance to the rules of high-level and low-level checklists

Dealing with Failure

A few words about failure

- Your project must have a non-negligible chance to fail
  - …else it would be development, not research
- Failure prompts the “I am not good enough” thought
  - Everyone (literally!) can do research. Work + perseverance make the difference
- Everyone fails, but beware of survivor bias + Friendship paradox
  - Only successes are made visible, and winners are loud about it
- Reach out!
  - Others will help you put things in perspective
- Dissect failure and learn from it
  - At which point of your framework failed? (e.g, Method? Question?)

Veronika Cheplygina “CV of Failures”, “How I Fail”
https://veronikach.com/failure/

Now go free and do your research!

Questions? Experiences to share?

Tldr; summary: https://wiki.itu.dk/computerscience/index.php/Research_Project_and_Thesis