

UX for AI

July, 2023

Table of contents

Introduction	3
The Basics	4
What is UX for AI?	4
Designing UX for AI is different	4
Responsibilities of an AI designer	6
A deep understanding of people, systems, and data	6
Prototype and test your AI with users early and often	7
Be ethical from the outset	8
Users of an AI system	9
The Elements of AI User Experience	10
AI Intent	11
Design for human and AI contexts	12
AI contexts	13
AI Modalities	16
AI UX components	18
UX for AI Design Principles	19
AI must be human-centered	19
Simple yet powerful	19
Design for human perception	20
Assistive automation	22
Earn trust with transparency	23
Always build for explainability	23
Designing for a human in the loop	24
Feedback is essential, not an afterthought	25
Informative insights from data	27
Design for resilience	28
Conclusion	30
Appendices	31
About the team	31
Learn more	32
Glossary	33
Resources and links	35

Introduction

We must influence and work towards building intuitive and effortless solutions with **artificial intelligence**, making people's lives easier by augmenting their capabilities. This body of work is a first step in updating and expanding our craft of user experience design to support this rapidly evolving space.

The goal of this document is to document baseline AI design guidance so that:

1. Designers can expand their craft knowledge to become competent AI designers at IBM.
2. Team collaborators can gain empathy and understanding of AI-UX design without needing to be deeply versed in the details of the craft.

We need to think of AI as something that interacts with us using what it has learned rather than just following strict rules. These new experiences change the nature of our design considerations and add a new level of complexity and responsibility to our work; our understanding and craft must evolve.

We invite you to collaborate with us to grow and nurture this work.

The Basics

What is UX for AI?

Put simply, **User Experience for AI** (UX for AI) drives everything a user sees, hears, feels, does, and achieves with AI-based applications.

UX for AI refers to designing interfaces and interactions between humans and intelligent systems to maximize user satisfaction, build trust, and achieve business goals.

If used responsibly, AI can augment us and make our work easier, faster, more trustworthy, and more reliable. As AI designers, we must ensure people can trust the systems we design and be aware of their potential consequences. We should also be open and honest about how our AI works.

AI is transforming how we interact with digital systems, from chatbots and voice assistants to intelligent recommendation systems and personalized services. However, designing effective and engaging digital user experiences for **AI-infused applications** requires unique skills and considerations that differ from traditional user experience design.

Designing UX for AI is different

Consider what we've learned living through the past few major technological innovations. The World Wide Web brought the world's information to our beck and call and taught us that whatever we want, whenever we want, it is always a finger-length away. Mobile computing enabled us to have that information with us at all times. Social media changed the foundations of how we communicate with each other and language itself.

Each innovation provided a new context for communication and expanded (or changed) our understanding of the relationships we have with machines. AI requires us to take notice of new contexts yet again. This time, designers must account for a system that can understand, reason, learn, and interact.

As a result, this changes the nature of our design considerations, some of which we take for granted based on the last 40+ years of software design. If we've moved past typing into a text field and pressing "submit," what does it mean to evolve our craft for artificial intelligence?

Data literacy is required

As designers, we're rarely, if ever, taught about the significance of data. Traditionally, being knowledgeable about data isn't considered part of our skillset. But it's changing in real-time thanks to the emergence of AI.

Data is the lifeblood of artificial intelligence. It's the fuel that powers the AI engine. In its simplest form, data is a collection of facts and statistics that we use for reference and analysis. It's a record of what has happened or is happening, whether it's graffiti, nutritional facts, or data visualizations of sales.

Data is an artifact of human behavior. It's a reflection of who we are and what we do. When we gather well-researched and curated data, we better understand ourselves and the world around us. We can make better decisions, gain insights, and even predict what might happen in the future.

So, if we're going to create truly impactful AI experiences, we must embrace the importance of data. The value of your AI relies heavily on the quality and understanding of your data. Once you learn how to harness the power of data, you can design experiences that aren't only data-driven but also human-centered.

AI is probabilistic

Throughout our design careers, we've primarily focused on **deterministic** applications — systems that consistently produce the same output for a given input. The behavior of these systems is entirely predictable, driven by explicit logic and input. However, we're now working with **probabilistic** AI, which makes decisions based on probabilities and statistical reasoning. Unlike deterministic applications, probabilistic AI acknowledges the inherent uncertainty and complexity of real-world problems. Employing probabilistic models and statistical inference offers a more nuanced understanding of phenomena.

Designing probabilistic AI systems requires a shift from rigid rules to working with probabilities. These systems continuously adapt and improve their decision-making capabilities by learning from new information. So, instead of aiming for definite results, we now have to design for what's most likely to happen. This means we need to be as flexible as possible to accommodate a wide range of possible outcomes. It's what we call "**generative variability**." This variability can make it challenging for users to achieve the same results repeatedly. However, generative applications have their perks. They allow users to explore different possibilities, experiment with new ideas, spark creativity, and gain knowledge in new areas.

As designers of AI systems, we need to adjust our standards to account for the probabilistic nature of these systems. We must provide clear explanations and supporting context to ensure that our designs are flexible enough to accommodate diverse outcomes.

Accounting for bias

Bias is a significant concern when it comes to AI. Whether conscious or unintentional, biases can seep into AI algorithms through the data or human inputs they learn from. Think of it like learning a new language from a single teacher with a strong accent - you might start speaking the same way. Similarly, if an AI algorithm is trained on biased data or limited perspectives, it can produce biased results, leading to potential discrimination or worse.

To tackle this issue, as AI designers, we must take on the role of user advocates. It's essential to identify and address biases in the data before building AI models, a responsibility that lies with **Data Scientists**. However, there is no one-size-fits-all solution to mitigate bias, and it requires careful adaptation and leveraging of the user experience. Furthermore, diversity and inclusion play a vital role. By having diverse teams collaborate in creating AI systems, we can ensure a broader perspective and minimize bias, ultimately designing AI experiences that benefit everyone and positively impact the world.

Responsibilities of an AI designer

The success of a user's experience with any solution, technical or otherwise, is directly correlated to how well the solution solves a problem. Hence, an AI designer must **prioritize practical problems to solve with AI**. It's not enough to ask, *"Is this possible?"* As always, we're responsible for asking, *"Should we do this?"* and *"Will this be useful?"*

As AI designers, we must **outline a compelling vision** that brings about an intersection of user needs, business goals, and technological capabilities in the right place at the right time to enable a beneficial experience with AI. We must also work with our cross-disciplinary teammates to **define an intentional structure of system behavior** to make that vision a reality. System flowcharts, experience maps, and journey maps are great tools to collaboratively build this structural scaffolding so the team can see how the various pieces will come together to make a satisfying whole.

As stated in [a Fast Company article](#), *"Ultimately, now as in the future, designers have three key roles: to create (structure knowledge and represent perceptions); translate (move between contexts, platforms, and cultures); and articulate (give clarity to thought and feeling)."*

AI designers must ensure that, minimally, the AI solution satisfies a core need of users and then strive to exceed users' expectations of how it can augment their ability to fulfill an objective.

A deep understanding of people, systems, and data

- **People:** We're inspired by how our users operate and what they think about. We're informed by what motivates, concerns, empowers, and disempowers them. We must live in the hearts and minds of our users. It's our job to be their best advocates based on their needs.
- **Data:** Data enables us to better understand ourselves and the world around us. Data allows us to make better decisions. Lastly, data is the fuel for insight and, more importantly, foresight. Without understanding how to leverage data properly, creating beneficial AI is very difficult, if not impossible. The power of data is leveraging the past for a better future. Used well, data can allow you to make strong predictions about what can or will happen.
- **Systems:** We understand and respect how AI systems are structured and built. While this doesn't mean we need to write code, we need foundational knowledge of the technology to make it understandable, clear, and satisfying. Perhaps most importantly, it helps us, as designers, to avoid suggesting "sci-fi solutions" that can't be built.

Prototype and test your AI with users early and often

By leading transformational change, we're designing new experiences every step of the way. We must actively seek to prototype and test these experiences early and often to learn their flaws and improve upon them. Prototypes can be tested with actual data or using a [“Wizard of Oz” technique](#) with fake data and/or make-believe intelligent interactions to test the quality of the experience. With recent advances in [Large Language Models](#) (LLMs), it is even easier and quicker than ever to build concepts at various fidelities – from rough visuals to functional prototypes.

When validating prototypes, strive to validate them across a diverse range of participants, seeking feedback from not just users of an AI system but also [makers](#) and [affected people](#) wherever applicable. This early and continual testing is necessary to learn about gaps in an AI-augmented experience, often uncovering important issues around [trust](#), [reliance](#), [explainability](#), [feedback mechanisms](#), and the overall usefulness of the AI solution.

When testing prototypes, consider these areas to focus your learning and validation goals:

- **Purpose:** Does the prototype address a business need and a user problem that's worth solving?
- **Value:** Do users and/or potentially affected people agree that the proposed solution provides them intrinsic benefits that make their efforts easier or more valuable?
- **Trust:** Can users of the AI system and those affected by it trust the system's rationale and recommendations? What factors are crucial in earning their trust? Unpacking what factors lead to more trust in an AI system and the output it generates will lead to more effective solutions.
- **Usable:** Can makers and users of the AI system perform their tasks successfully in the prototype? Is the mental model clear? Do they understand how to make use of this system as designed?

- **Interpretability:** Can users of the AI system correctly interpret the outcomes of the AI system? Do they understand the output being generated by the system as it is articulated? Do they understand if there are any pros and cons associated with this output? Similarly, can people affected by the AI system formulate a possible mental model of the AI outcome?
- **Task efficiency:** Can the users of the AI system perform their jobs more efficiently or with greater success when using this tool compared to when not using it? If yes, is it a considerable increase in efficiency?
- **Possibility of misuse or misconception:** Suppose you didn't include guardrails or context of use with the AI system, and people could use it as they wish. How do users or people potentially affected by it imagine its usage? Can any of those usage scenarios lead to undesirable outcomes? If yes, what guardrails could help you prevent or mitigate such unchecked use of this system?

“AI systems are undeniably powerful tools. And like all powerful tools, great care must be taken in their development and deployment. The first step in this process is to build systems that can be trusted. This will require a framework of best practices that incorporates appropriate values and ensures ethical behavior, including alignment with social norms and contracts, algorithmic responsibility, explanation capabilities, compliance with existing legislation and policy, assurance of the integrity of the data, algorithms and systems, and protection of privacy and personal information.”

Francesca Rossi

IBM Fellow and AI Ethics Global Leader
IBM Research

Be ethical from the outset

Ethical decision-making isn't another form of technical problem-solving. AI teams too often fail to thoroughly explore the ethical implications of their day-to-day responsibilities. AI designers must respect human impacts for any AI solution. As AI designers, we're responsible for leading our users to a capable understanding of our AI tools.

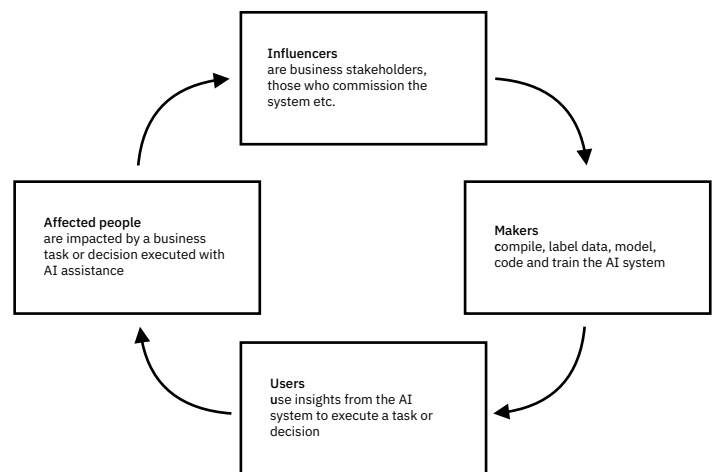
- **Team alignment:** We need a comprehensive understanding of AI systems' real-world effects and outcomes to know where, why, and how to integrate ethical guidance into our work. Sometimes, technical practitioners don't interact enough with end-users to understand these impacts. As designers, we can lead conversations and [group exercises](#) to better forecast these impacts for any new product or feature we want to create.
- **Encouraging ethical behaviors:** A [core skill](#) of an AI Designer is emulating ethical behaviors and practices on their team. Designers should be able to demonstrate skills such as telling real-world stories about the human impacts of unethical design decisions, asking key questions to uncover risks and areas of bias, and integrating ethics resources into existing processes.
- **Explainable and transparent design:** AI systems should be designed with transparency and explainability mechanisms in the UI. This requires a deeper understanding of users' AI mental models, users' explainability needs, compliance requirements, and an ability to leverage any existing patterns that improve user understanding. Translating the workings of an AI system is necessary to building trust over time.

Users of an AI system

In addition to the conventional classification and understanding of computer system users, AI systems require you to consider AI-specific user archetypes. The diagram below shows the four main archetypes of the users of the AI system and where they fit within the AI system cycle.

- **Influencers** such as stakeholders are involved in commissioning the AI system based on user needs, and others must generate the data needed to train the AI system.
- **Makers** then compile, model, code, and label the data to train the AI system.
- **Users** use the decisions and insights generated from the AI system to make decisions or execute a task, and these decisions or tasks affect other people.
- **Affected people** are directly or indirectly impacted by the outcomes of AI-assisted decisions.

For example, a roadway near a school has recently seen an increase in speeders at certain times of the day. Policy analysts (**influencers**) commission an AI system based on driver data. Data Scientists (**makers**) compile this data, then code and model it to train the AI to determine speeding instances based on driver speed and the time of day. Police officers (**users**) use the insights generated by the AI to decide if a driver should receive a speeding ticket. The AI's decisions impact drivers (**affected people**), and they may receive a speeding ticket.









These user archetypes inform how we create and manage useful AI experiences for humans interacting with them – from determining how and when to bring a [human into the loop](#) alongside the AI system or figuring out how and when to seek feedback to course-correct or to know how best to provide value through the AI system.

The Elements of AI User Experience

The UX of any AI consists of these six layers. They are unchanging and consistent, regardless of how an AI manifests.

As AI Designers, it is our responsibility to understand the importance and requirements for each of these layers. Perhaps more importantly, it's essential for AI Designers to understand the interplay between the layers to craft the best possible experience for our users.

 AI Intent	 AI Ethics	AI Intent is the fusion of business needs and user needs. Business Needs capture what the business requires to succeed, User Needs illuminate what the end user is trying to achieve.
 Human Contexts		Human Contexts are the emotional, physical, and social dimensions that shape our experiences and interactions with the world.
 AI Contexts		AI Contexts provide context for the system and its users. Context determines an idea's relationship to another idea. In the case of an AI design concept, a context is the mode in which an AI solution operates.
 AI Modalities		AI Modalities refer to how people perceive, understand, and interact with artificial intelligence. Primarily, this refers to how users navigate and interact with AI.
 AI UX Components		AI UX Components are comprised of small, reusable interface elements that can be combined and juxtaposed to create more complex user interfaces.

AI Ethics are the principles, guidelines, and practices that ensure responsible and accountable development, deployment, and use of AI, with a focus on addressing potential ethical, social, and legal implications.

AI Intent

Understanding User and Business Needs

AI Intent is the fusion of business needs and user needs. Business Needs capture what the business requires to succeed, User Needs illuminate what the end user is trying to achieve.

As Charles Eames wisely stated, *“Recognizing the need is the primary condition for design.”*

Design should never be a result of chance or accident but rather the deliberate application of intent. It should address the specific requirements of both the user and the business, ensuring that the solution aligns with their goals and effectively solves the problem at hand.

The significance of user needs cannot be overstated. The purpose of any system, especially with AI, is to serve its users. The intent behind the design should stem from a deep understanding of what the user wants and needs. We can only create solutions that resonate and provide meaningful, lasting value by aligning with user needs.

At the same time, we can’t overlook the importance of business needs. Clear alignment with business goals is critical for the success of any effort. Whole-team alignment is essential to achieve successful outcomes. Every team member must have a shared understanding of a well-articulated intent driving the design and build of an AI. Rushing into implementation without a clear idea of what success looks like for both the business and the users can lead to suboptimal results. Alignment on intent from the start allows teams to work towards a common purpose, enabling efficient and accurate execution.

To design AI solutions with purpose and intention, a set of guiding principles can be adopted. These principles include:

- **Relevance**, ensuring the design solves real problems and aligns with user and business goals
- **Human-centeredness**, considering human needs, limitations, and behavior in the design process;
- **Transparency**, clearly communicating complex system dynamics; and responsibility, considering the broader social context and addressing issues such as fairness, bias, and unintended effects.

We must prioritize defining and articulating business intents *before* diving into technical implementation. This approach ensures that teams understand what success looks like for a given effort and allows for more effective execution. AI Intent also requires it to be written down and available to *all* team members at any time. This gives teams the ability to reflect, debate, and align. This alignment is the critical outcome provided by an AI Intent.

Adopting an approach to AI design that prioritizes human users and aligns with business needs is essential. This approach spans all phases of development, from understanding user needs and problem formulation to the final delivery of results. By placing user needs and business goals at the forefront, designers can create AI solutions that are not only technically viable but also valuable, responsible, and transformative.

Articulating clear user and business intent is key to everything we do.

Design for human and AI contexts

Context determines an idea's relationship to another idea. In the case of an AI design concept, a context is the mode in which an AI operates. To be consciously aware of the primary context affords designers the ability to intentionally dial in certain traits to more accurately deliver cognitively-derived results.

To design useful and well-aligned experiences with AI, we need to understand how humans function and behave in various contexts and how AI can function across its own contexts. This understanding is the basis for envisioning a fruitful collaboration of humans and AI tech toward any beneficial outcome.

As outlined below, an AI experience must adhere to human contexts and expectations while staying true to the underlying machine's projected context in the form of the role it's supposed to play and how it engages with humans.

Human contexts

Human contexts are comprised of emotional, physical, and social dimensions that shape our experiences and interactions with the world.

Emotional context is the range of human emotions, subjective experiences, and affective states we all live with, and the impact this has on individuals' well-being and overall user experience.

Physical context encompasses the tangible and sensory elements of our environment, including the immediate physical surroundings, which impact our daily lives and interactions.

Social context revolves around group norms, dynamics, and intimacy levels, influencing behavior and communication patterns and shaping the formation and maintenance of relationships within specific situations.

Together, these contexts intertwine to create a complex web that influences our thoughts, feelings, and actions, ultimately shaping our overall human experience. By understanding these contexts, we can make choices that will help us better align our AI efforts.

Emotional	Physical	Social
<p>Emotional context is the range of human emotions, subjective experiences, and affective states we all live with, acknowledging the impact they have on individuals' well-being and overall user experience.</p> <div><div>Happy ↔ Sad</div><div>Excited ↔ Anxious</div><div>Serene ↔ Stressed</div><div>Flow ↔ Frustrated</div><div>Assured ↔ Scared</div></div>	<p>Physical context refers to the tangible and sensory aspects of our surroundings and their impact on our daily lives. It includes the immediate physical environment and its influence on our interactions.</p> <div><div>Indoor ↔ Outdoor</div><div>Quiet ↔ Loud</div><div>Passive ↔ Active</div><div>Seated ↔ Mobile</div><div>Public ↔ Secure</div><div>Comfort ↔ Distress</div></div>	<p>Social context involves group norms, dynamics, and intimacy levels, shaping behavior and communication patterns. It influences how we form and maintain relationships within a given situation.</p> <div><div>Alone ↔ With others</div><div>Engaged ↔ Detached</div><div>Formal ↔ Informal</div><div>Funny ↔ Serious</div><div>Public ↔ Private</div></div>

AI contexts

Just as we humans have our own contexts through which we experience the world around us, AI has its own contexts that need to be considered when designing it. An AI's context is the role AI plays in the human + AI collaboration.

While it's not entirely accurate, you can think of an AI context as a high-level persona for an AI.

Note: These contexts are archetypes. While there will be many instances where they manifest as written below, the reality is that most AI design efforts will blend contexts by mixing and matching relevant characteristics based on the needs of the solution and the users. The point here is to give designers and product managers a shared vocabulary with shared understanding of the desired outcomes.

Contexts and mental models

[Research in human-AI interaction](#) suggests that in forming a mental model of an AI application, users are likely to ascribe context to it. **Users' expectations and actions will differ based on which context they're interacting with.** Clearly establishing the context of an AI application in a user's workflow, as well as its level of autonomy, will help them better understand how to interact effectively with it.

When determining the context of an AI, consider the following:

- Is it a tool, a partner, a teacher, a coach, or an assistant?
- Does it initiate actions or does it just respond to the user? (In other words, does the AI have [agency](#)?)
- Does it change an artifact or process directly or does it make recommendations/surface insights to the user?

Assistant

An Assistant is reliable, efficient, and supportive. It's there to assist the user when asked, therefore placing the bulk of the interaction in the hands of the user. The Assistant context delivers AI capabilities in a toggle mode where it is consciously beckoned forward by the user. A variant of this context has agency and is able to call itself forward when sensing a user's need.

Key characteristics:

1. **Attention to detail:** An Assistant should be detail-oriented, ensuring that tasks are identified and completed accurately and to a high standard.
2. **Prioritization skills:** An Assistant should be able to prioritize assigned tasks effectively.
3. **Communication skills:** An Assistant should have excellent communication skills to ensure that messages are conveyed clearly and accurately.
4. **Initiative:** An Assistant should take initiative to identify and complete tasks that need to be done without being asked...when applicable and desired.
5. **Adaptive:** An Assistant should be able to adapt to changing priorities and requirements.

Coach

The Coach is there to guide users within the actions they take to help guide them to a better outcome. A good Coach inspires and guides users toward achieving their goals. It takes the lead confidently but knows just when to step aside to let the user make their own judgement as to how to proceed. Its presence is noticeable but not the sole focus.

Key characteristics:

1. **Knowledge and expertise:** The Coach should have a deep understanding of the subject matter it's coaching on and possess the relevant skills and experience to provide effective guidance.
2. **Communication capabilities:** A Coach should be able to convey information clearly and concisely while actively listening to its users.
3. **Supportive:** While no machine is capable of having actual EQ, a Coach should be able to provide a supportive and non-judgmental guidance.
4. **Adaptability:** A Coach should be able to adapt its coaching style and approach to meet their users' needs and learning styles.
5. **Goal-oriented:** A Coach should be focused on helping its users achieve their goals, providing clear guidance and feedback to ensure progress is being made.
6. **Accountability:** A good coach should set clear expectations and help users stay on track toward their goals.

Teacher

This context is straightforward, the Teacher is there to educate the user. In this role, the AI shows the user how to get something done properly and helps to connect the dots so the user can learn new skills and information that will enhance their jobs. A Teacher creates a positive and supportive learning environment.

Key characteristics:

1. **Knowledge and expertise:** A Teacher should have a deep understanding of the subject matter and possess the relevant skills and experience to provide practical guidance.
2. **Approachable:** An effective Teacher is approachable and creates a safe and comfortable learning environment, which can encourage students to ask questions and seek help when needed.
3. **Communicative:** A Teacher should be able to convey information clearly and engagingly, especially in terms of its ability to explain complex concepts.
4. **Adaptability:** A Teacher should be able to adapt its teaching style and approach to meet its users' needs and learning styles. It also uses a variety of teaching methods to engage users and promote understanding.
5. **Accountability:** A Teacher should hold its users accountable, setting clear expectations for progression and providing constructive feedback to help them improve.
6. **Encouraging:** A Teacher should be encouraging and supportive, helping users to build confidence and believe in their abilities.

Partner

The Partner context is there to share the experience with the user as it progresses. It learns with the user and highlights what it considers to be relevant as if it were co-experiencing the solution with them. This is best thought of in terms of Pair Programming, a software development approach where two developers sit together working on one computer. When applied to an AI, the AI plays the role of the other person. It's there to reason through a problem space in tandem with the human but not just centered around code but on anything the AI is trained to do. A Partner works successfully *with* its users towards a common goal.

Key characteristics

1. **Communication skills:** A Partner is trained to ensure its messages are conveyed clearly and accurately. It pays attention to users' input and responds appropriately.
2. **Flexibility:** A Partner should be open to new ideas and approaches, demonstrating a willingness to adapt and change course when necessary.
3. **Problem-solving skills:** A Partner has strong problem-solving skills, working collaboratively with its users to find solutions and overcome challenges.
4. **Goal-oriented:** Partners are focused on the end goal, working with users to achieve a common objective.

AI modalities

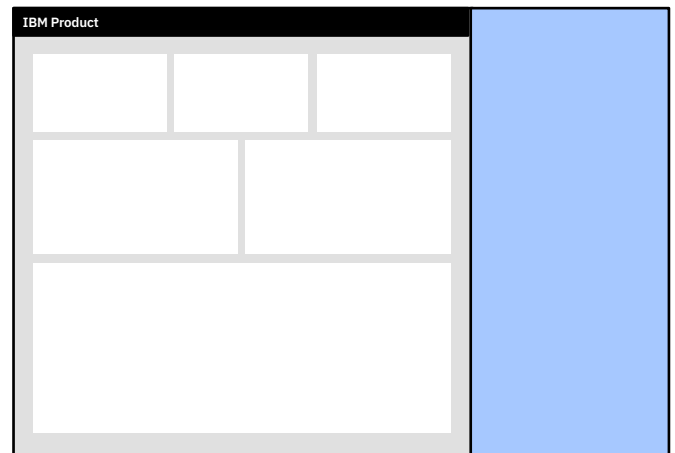
AI modalities refer to how AI presents itself to its users. In other words, it's how users perceive, understand, and interact with an AI. Primarily this refers to how users identify and navigate using AI, including helping users understand what they're doing and whether they're on the right track.

By considering how to apply these different modalities, designers and developers can create experiences that are functional, easy to use, and enjoyable for users. The goal of AI modalities is to create seamless and satisfying experiences for users, which leads to increased engagement and satisfaction with our products and services.

Sidebar modality

The sidebar:

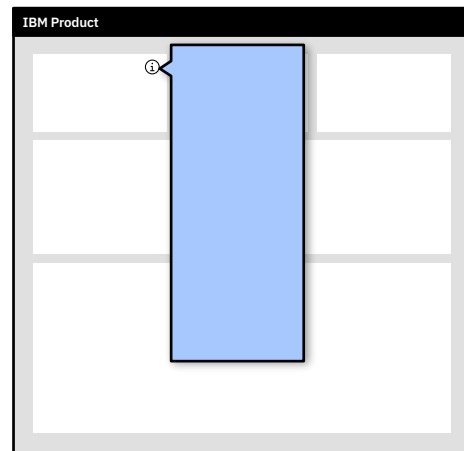
- Slides out or is omnipresent
- Can contain conversational UI, buttons, graphs, tiles, cards, text, images, or a mix.
- Primarily leveraged for deep interaction with main content



Call-out modality

The call-out:

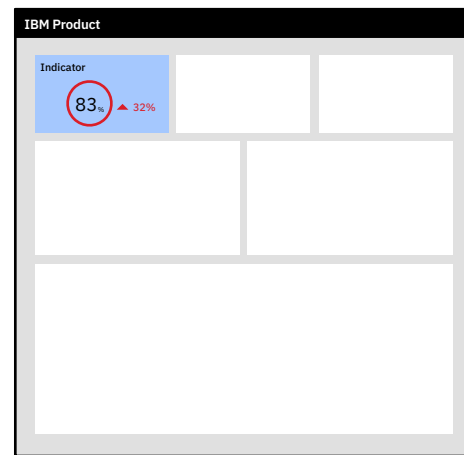
- Is used to “peer under the covers” to get a sense of the AI’s archetype in presenting this information
- May include lightweight functionality through a button and/or feedback loop
- Invoked through click or hover



Ambient modality

The ambient modality:

- Is a sibling to the call-out modality
- Displays results, outcomes, or insights generated by an AI
- Is limited in interactivity, more for display purposes (think speedometer/tachometer)



Spotlight modality

The Spotlight modality:

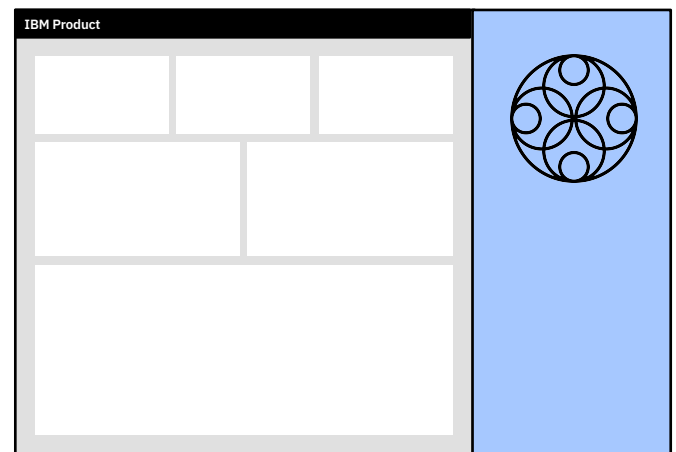
- Puts the AI-driven experience front and center
- It is the focal point of a user's interaction



Avatar modality

The Avatar modality:

- Is typically text or voice interaction only using NLP/NLU
- Should *never* attempt humanoid form. Should always be abstract.
- Likely has some degree of subtle animation



Hinting/completion modality

The hinting/completion modality:

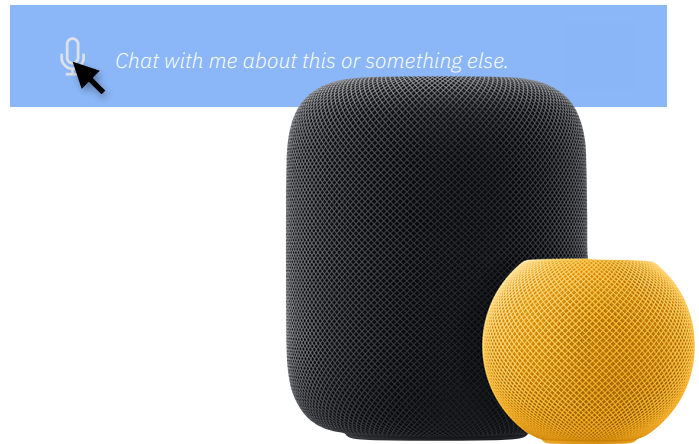
- Is generally text/ASCII-based
- Suggests what should come next given a specific prompt
- May hint or complete 1 line or an entire body of code/text

```
11
12 - name: Add new user
13   ibm.ibm_zos_core.zos_tso_command:
14     commands: "ADDUSER {{ userid | upper }}"
15               DFLTGRP('{{ default_group | upper }}')
16               AUTHORITY('{{ default_group_authority | upper }}')
17               OWNER('{{ owner | upper }}')
18               NAME('{{ name | upper }}')
19               PASSWORD('{{ password | upper }}')
20               PHRASE('{{ passphrase | upper }}')
21               SECLABEL('{{ security_label | upper }}')
22               SECLEVEL('{{ security_level | upper }}')
23               ADDCATEGORY('{{ category | upper }}')
24               TSO(ACCTNUM('{{ tso_account_number | upper }}') PROC('{{ tso_logon_procedure | upper }}'))
25               DFP(DATAAPPL('{{ dfp_data_application | upper }}') DATACLAS('{{ data_class | upper }}') MGMTCLAS('{{ management | upper }}'))
26               OMVS(UID('{{ omvs_uid | upper }}'))
27               HOME('{{ omvs_home_directory | upper }}')
```

Audio-only modality

The audio-only modality:

- Is voice interaction-only using NLP/NLU
- May be invoked with a mouse click when embedded in an on-screen experience instead of a smart-speaker



AI UX components

As AI is increasingly integrated across all modalities of our software, transparency of AI is a necessity to establish and maintain trust with users.

The Carbon for AI component library is an extension of the Carbon Design System intended to provide a visually and behaviorally distinct identity for any instances of AI within our products. It consists of a set of reusable design components and principles to establish a consistent understanding and representation of AI across our products.

Carbon for AI is an extension of existing component libraries that emphasize AI transparency and explainability and a set of guidelines and considerations when incorporating AI into new and existing experiences.

Carbon for AI is currently available as an alpha Figma kit. The Carbon coding is underway and expected to be complete by mid-February, 2024.

Access the [Carbon for AI Essentials Kit](https://www.figma.com/file/4aMIX5zjux2LhbW1aI3xgD/Carbon-for-AI---AI-Essentials-Kit?type=design&node-id=1971%3A23010&mode=design&t=in5r5AdLG3HhLvR3-1) here:

<https://www.figma.com/file/4aMIX5zjux2LhbW1aI3xgD/Carbon-for-AI---AI-Essentials-Kit?type=design&node-id=1971%3A23010&mode=design&t=in5r5AdLG3HhLvR3-1>

UX for AI

Design Principles

AI must be human-centered

Designing AI that works well for humans involves considering several factors to ensure that the AI system can perform its intended tasks effectively and efficiently while being easy for humans to use and understand. Some key considerations in designing AI for effective human use include:

1. **User needs and goals:** It's essential to clearly understand the needs and goals of the users who will be interacting with the AI system and to design the system in a way that meets those needs and helps users achieve their goals.
2. **User experience:** The AI system should be designed with the user experience in mind, considering factors such as the user interface, the speed and accuracy of the system, and the overall simplicity and ease of use.
3. **Human-machine interaction:** The AI system should be designed to interact with humans naturally and intuitively, using clear and concise language and providing appropriate feedback to the user.
4. **Safety and reliability:** Ensuring the safety and reliability of the AI system is critical, as it will interact with and potentially make decisions for humans. This includes testing, validating, and ensuring the system can handle edge cases and unexpected situations.

Ensuring that all of these factors are taken into account can help to create AI systems that are effective, efficient, safe, and easy for humans to use.

Simple yet powerful

Simplicity is the quality or condition of being easy to understand or do. It's a mistake to think simplicity means having a small feature set. Elegant, simple design contains the necessary, essential, and occasional features in a way that doesn't influence any of their uses, revealing and hiding them as necessary.

One of the most challenging aspects of AI design is keeping features out. Simplicity can be, in itself, a market differentiator. Our AI solutions can be made more elegant by simply omitting extra things. Be smart about what you don't do.

Design for the simplest, most common use path. Use progressive disclosure techniques and contextual relevance to handle exception cases so the experience maintains a feeling of simplicity for most users.

Some tenets of simplicity

- The details are the design. Simplicity isn't a matter of dumbing things down. Instead, it's about thoughtfully curating and managing details so the user doesn't have to spend time figuring out what's going on.
- Be transparent with your goals for your users. People like to have a mental model of how things work.
- Edit ruthlessly. Simplicity hinges as much on cutting non-essential features as on adding helpful ones.

Design for human perception

Human perception is how our brains interpret and make sense of the world around us by classifying, organizing, identifying, and interpreting sensory information. It's how we build an understanding and mental model of the world around us that constantly evolves as we update our models based on feedback from new experiences and lessons learned from the past.

For example, we're all taught to not put our hands in fire because it'll hurt. If you decide to test this, you'll likely try it once, get burned, and experience pain (feedback). You'll probably never repeat the test, as you've learned through experience and created a mental model in which extreme heat cause personal harm to bare skin.

Designing software experiences to cater to human perception is a complicated task, particularly when using artificial intelligence, given the probabilistic nature of these systems. We now have to account for how each user will identify, organize, and interpret the varying information supplied by the AI.

Consider an LLM that's generated with billions of input and processing parameters. This complexity poses a challenge as it is nearly impossible to test and design intentionally for probable output scenarios across various combinations of these parameters. To add to that, consider the various applications that LLMs facilitate – from specific usage like customer support solutions in a particular domain to general use like producing art or suggesting code blocks. This wide range of usage further complicates how humans perceive and resonate with such AI systems.

AI Designers must seek to understand human perception and how it works across an emotional range to create AI systems that are both effective and emotionally resonant for users.

Balance predictability and serendipity

Depending on the context of use, AI can be leveraged as:

1. A tool to scale human abilities in ways that are familiar and known to humans (i.e., predictable outcomes)
2. Or as a tool that leads to new, creative ways of thinking beyond commonly known human patterns of interpretation (i.e., serendipitous outcomes).

Predictable outcomes adhere to existing mental models of humans, and foster trust and confidence in the system, through familiarity.

Serendipitous outcomes push the limits of human thinking beyond what's known. They foster creativity through exploration.

As AI designers, we understand and define an intended range of experiences to support in our AI experiences, along this spectrum of predictability to serendipity. This understanding can help us determine if our experience should offer ways for users to build new mental models or if it should adhere to existing mental models.

AI Usage Spectrum



More predictable outcomes

These outcomes feel "probable" and "plausible," in line with current human mental models of the world.



More serendipitous outcomes

These outcomes feel "creative" and "inspirational," appearing to push the limits of possibility, beyond current human mental models of the world

Consider the **intent** of your AI application.

1. Does it enable innovation or discovery of new, previously unknown patterns? Is it about exploring new scientific research hypotheses, or making art in ways we previously didn't consider.
2. Does it enable a faster, more efficient way to complete a job the user already knows how to do? For example, quality control using specified procedures, inventory management, or any effort with workloads too larger or complex for humans to handle alone.
3. Or does it bring a mix of both intents, like creative problem-solving to spot hidden patterns? For example, human users may use AI to help them spot emerging, uncommon fraud patterns in financial data.

While the first case above might lean more towards establishing new mental models, the other two cases might require building upon familiar and predictable reasoning to take users from their current mental model of the situation and show them a new perspective.

Set expectations on usage

It's essential to set expectations with users right at the onset about what an AI system is capable of, its limitations, and its intended purpose. If we don't communicate these boundaries, it's possible to lead users to disappointment. Being transparent with boundaries not only enables us to form more accurate mental models of AI systems but also encourages adaptive collaboration, where the users can communicate if a system isn't aligned with their perceptions and as a result, needs to be course-corrected, or enhanced to better match users' expectations.

In addition to intent and limitations, it is also necessary to ensure clear and understandable interaction mechanisms without leaving users guessing about how to interact with an AI system. Unexpected triggers to activate an AI system or uninitiated responses from the system can detract from a user's experience and reduce confidence in the system due to a perceived lack of control.

For example, imagine a situation where an AI system is hallucinating, and the user is confused about what led to this, all because the system limitations that possibly caused this behavior weren't communicated. Or, consider an embedded AI system installed in a public environment that can record audio or visual imagery of the surrounding area. Without a perceptible presence indicator or ways to engage with this kind of system,

people may feel a lack of control leading to distrust, even if the system was initially intended to help them. The user experience in each of these scenarios could have been more fit for purpose, with some expectations provided at the onset about the system's presence, capabilities, and/or limitations.

A strong understanding of the [principles of emotional design](#) can help with cultivating a positive human-AI relationship, which is in tune with and respectful of human needs and expectations over time.

Personalize, don't personify

Personalization is a powerful tool for designing experiences that foster meaningful connections between users and machines. When an AI appears to "know" what we want, it can create a delightful user experience. However, it's essential to strike a balance between personalization and personification. While personification may seem like a way to make a warm connection with users, it can also lead to misguided expectations and misplaced trust in AI. Therefore, preserving the integrity of the AI experience while making it engaging is essential.

When crafting solutions, designers should focus on the details that contribute to personalization while avoiding the pitfalls of personification. A good way to achieve this is by using data to personalize the experience based on user behavior and preferences rather than relying on anthropomorphic characteristics. This approach ensures that users feel seen and understood without having to project human attributes onto machines. However, data-driven personalization should be balanced carefully with the potential for bias and unintentionally limiting users from exposure to different kinds of information. For example, a smart search that considers past queries from a user or similar users to provide suggestions can have a tendency to amplify or repeat confirmation, anchoring, or availability biases.

It's also vital to avoid sacrificing credibility for the sake of being overly witty or personable. While humor and personality can be effective tools for engaging users, they should be used judiciously and only when appropriate. Overuse of these elements can undermine the credibility of the AI system, leading users to doubt its effectiveness.

Assistive automation

When defining automation use cases, it's important for AI to provide assistive automation rather than over-automating tasks, to balance AI's benefits with the importance of oversight, ethics, and creativity.

Keep the user in control

We believe in using AI to augment and enhance human capabilities, not to replace human skills. To that end, we must ensure that our AI solutions work to complement what the user is doing and not supersede with control. Always allow a user to override a recommendation easily or to intervene and take over whenever they see fit. The user should be in control of any decision-making that involves subjectivity, with the AI working together with the user to guide and inform that decision. Remember, AI is a tool for a job to be done. It's not the job itself.

Automate only if it adds value

"Should AI do this task?" is a question we must ask ourselves every step of the way. Automation is the answer only when a compelling reason calls for it. Consider the following questions to help understand if automation would be valuable or not for a task, and if so, what would be necessary guard rails around it.

1. Which specific tasks within the wider workflow are being considered for automation?
2. What is the accuracy, speed, or efficiency with which a human can do these tasks compared to an automated flow? And how does this compare to an augmented human + automation flow?
3. What are the consequences of failure or a mistake being made by the automated flow? What happens if the automation doesn't take action when it was supposed to or takes a wrong action? What might be a factor leading to a wrong action?
4. Based on the consequences of failure or errors, what kind of human oversight is required for each automation task?
5. Should human oversight be proactive or reactive? What is the necessary reaction time? How will the automation respond to this reaction? How do we minimize over-reaction from the automation or human counterpart?

A comprehensive outlook guided by these attributes will help assess the effort and value of automation, eventually determining whether this task or the broader workflow is a suitable candidate for automation.

Weigh the benefits vs risks of automation

- Can automation minimize human oversight in a way that leads to a lack of clear accountability? This is particularly problematic in high-risk use cases where the consequences of bad decisions can be significant.
- Can automation lead to ethical or moral issues? For example, how would potentially biased decisions from the automated system be identified or mitigated without leading to unfair or unethical outcomes?
- Can automation lead to a lack of creativity or innovation in problem-solving? For example, suppose an AI system can fully automate tasks with a certain level of subjectivity. In that case, it can lead to repetitively similar outputs devoid of unique human interpretation or taste.

Levels of computer automation
Sheridan & Verplank, 1978

Complete autonomy	▲ 10. The computer decides everything, acts autonomously, ignoring the human.
Some automation	9. The computer informs the human only if it, the computer, decides to.
	8. The computer informs the human only if asked.
	7. The computer executes automatically, then necessarily informs the human.
Low assistance	6. The computer allows the human a restricted time to veto before automatic execution.
	5. The computer executes that suggestion if the human approves.
	4. The computer suggests one alternative.
	3. The computer narrows the selection down to a few alternatives.
No assistance	2. The computer offers a complete set of decision/action alternatives.
	— 1. The computer offers no assistance; the human must take all decisions and actions

Earn trust with transparency

Humans are wired to be wary of new things, situations, or even new people, approaching them cautiously. To consider something or someone “trustworthy,” we tend to look for consistent behavior indicators. We want to know when to expect a response and when not to expect it. We want to know when they’re reasonably knowledgeable on a topic and when they aren’t. We want to know when they make a mistake. We want to see that they are candid, truthful, and reliable. This behavior extends to how we perceive and assess emerging technologies like AI.

Any AI experience must gain users’ trust over time by showing a consistent pattern of value and candid transparency. As builders of AI, we can intentionally design for trust by mapping an entire user journey through our experience, identifying low points where their trust is likely to break down, and understanding what might lead to such a breakdown. Then, we can use this detail to intentionally craft a “peak” moment of trust and delight in the journey, exceeding their expectations with transparency and candid behavior.

Consider these questions to help identify low points of trust in a user’s journey:

- Are you asking the user to take a leap of faith with recommendations or suggestions from the AI system? If yes, identify those touch points and unpack what information could transform that “leap of faith” into a “bridge of trust” to help a gradual move for the user.
- Is there any scope for your AI system to generate creative content based on its general understanding of human language or the user’s task or problem domain? If yes, what are the implications if the system hallucinates and produces content without regard for its meaning? Identify such places in your system journey where you rely on the system to produce useful content and ensure that a user can backtrack the system from undesirable hallucinations if needed.
- Is your system producing output toward a high-stakes task? If yes, what would enable the user to ensure that any output is high quality and to double-check that the system did not make a mistake and drop the ball?
- Provide explainable context to the user so they can see key assumptions or interpretations made by the system and question them as needed.

Always build for explainability

Explainability is a requirement for building healthy AI experiences. For humans to use AI to augment their decision-making, they need to understand how AI is built, deployed, and operated. As bonafide user champions of their teams, designers must prioritize explainable and transparent AI in every stage of design and development. As outlined in the [AI Explainability primer](#), we can explain particular model recommendations or predictions, broader training data, performance, system functions, and other model details.

Explainability is surfaced based on the user’s context, needs, and goals. You must consider the knowledge your user brings to the experience and the right balance of information they need to be successful and foster appropriate trust in the AI system. In some cases, this means the user doesn’t need much detail about its inner workings, but they should be able to learn more if required or preferred.

Explainability also depends on the type of AI system you’re working with. Your AI system might be able to generate explainability in the form of weighted data points or confidence scoring. You might use charts or graphs to visualize how the model compared data points. You might use tooltips to define key terms and provide links to more detailed information or use documentation to give the full background of the model, its training, deployment, and so on. Your model might be a black box, in which case any explainability you provide approximates the model’s behavior. In those cases, you can also be transparent about how the system has been designed and developed.

Regardless of your model type and purpose, you can leverage explainability in your UI experience in different ways. The AI might be ‘right’ in its reasoning, but accuracy and performance in AI are not enough to build trust and understanding.

Designing for a human in the loop

Human in the loop (HiTL) in AI refers to...

1. A framework of structuring general human-AI decision-making interactions, where the human is informed by the AI, but has the final say.
2. A framework specific to data science that describes how best to help a data scientist train a new data model. (Not discussed here.)
3. (As we will make the case below) an ethical priority.

How is HiTL different

HiTL sits alongside other models that distinguish the degree of agency that the human and AI have with regard to decision-making.

- In an *algorithm-in-the-loop* model (AiTL), the human user has the option to request AI information. Most modern productivity tools are like this, with AI features available on request. Think PowerPoint and its CoPilot. You can use PowerPoint in “manual” mode without touching the AI at all.
- In a *human-on-the-loop* model (HoTL), the AI will take action in a given time unless a human says otherwise. A Roomba vacuum cleaner is a good example. Once you set it up, it will do what you’ve told it to do as you told it to do it until otherwise or it runs into problems. [Christopher Noessel has written design guidance](#) for this under the name “agentive technology.”
- In *automation* (which can be thought of as *humans-out-of-the-loop*), the AI makes decisions and takes action until stopped by a human. An AI algorithm that optimizes packet-switching flow would be an example of this. No human wants to be in that loop, except perhaps to occasionally monitor its performance.

The main distinguishing feature for HiTL is that while AI inputs are a default part of the workflow, humans are the gate for final decisions. In this model, the AI cannot act without human initiative or approval. One of us has to press that button.

AiTL and HiTL will probably be the most common models designers can expect to encounter, for the foreseeable future.

Designing HiTL

Strictly speaking, all you need to do to design an HiTL interaction is ensure that the human has the final say. Users must be free to disagree with an AI recommendation, and the workflow of disagreement must not be so onerous that it encourages agreement-by-default.

Refer to the best practices for *explainability* so users understand the AI’s recommendation well enough to properly evaluate it, and the *Chiron* project to follow best practices for encouraging the right level of reliance.

The practical benefits of HiTL over other frameworks

In his book *Human in the Loop Machine Learning: Active Learning and annotation for human-centered AI* (Manning, 2021), author Robert Monarch notes that AI systems perform better with human feedback. Humans improve the accuracy of models. Ideally we reduce errors in data. We lower the risk of costly AI mistakes. And in data science, shipping models faster. So there are lots of reasons to consider this *first* when structuring an AI interaction.

But moreover, we see HiTL as an ethical imperative.

HiTL as an ethical imperative

We believe designers should consider HiTL models first, as the default, across all stages of an AI system. Adopt one of the other models only when HiTL does not fit. From how AI is built to how it is used every day, having a human oversee the system helps ensure that humans are aware of what the AI is doing (transparent), we can raise the alarm if it’s recommendations are troubling (accountable), and we can make the right decision in edge cases (ethical). Without human oversight, AI systems can drift into bias and amplify the negative consequences by re-learning the same biased signals it was trained on.

Designers should consider that the loop they’re designing for isn’t limited to the user. It should include any affected individuals subject to an AI-assisted decision. They should be made aware how decisions were made, and can easily request review and reconsideration.

And though it’s a matter for system architects, designers should specify in their documentation and discussions that feedback from its users does get fed back into the model(s) in ways that improve the model’s accuracy, fairness, and relevancy.

Feedback is essential, not an afterthought

Feedback is an essential aspect of AI systems because it allows for course correction and ensures the system is aligned with desired user values and outcomes. Without feedback, an AI system may continue to operate in a way that's misaligned with these values and outcomes, potentially leading to negative consequences.

Feedback can be initiated by either humans or the system itself, depending on the context and the stage of the AI lifecycle. For example, "thumbs up/down," "like/dislike," or "accept/reject" are examples of human-initiated feedback actions. Whereas system-initiated feedback actions may come in the form of specifically designed feedback prompts like "Was this helpful?" or "Was this relevant?" that launch a focused workflow to capture user feedback.

Whether the feedback is human-initiated or system-initiated, it must be integrated into the AI to allow ongoing course correction and improvement. This can involve incorporating feedback mechanisms into the system's design and regularly reviewing and adjusting the system's performance based on feedback data.

Overall, treating feedback as a core tenet rather than an afterthought is important because it helps ensure that an AI system is aligned with desired user values and outcomes and can be adjusted to continue meeting these goals.

Feedback on impact

We must seek to understand and qualify how an AI system could adversely impact the lives of people who are affected by AI-assisted or recommended decisions. Consider these questions to assess the possible impact of such decisions on them:

1. Could imperfect or inaccurate AI recommendations lead to life-altering consequences?
2. How can we mitigate such high-impact consequences?

Mitigation mechanisms can be initiated by a human or by the system itself and can be broadly considered in 2 categories:

1. **reactive mitigation:** happens after the decision event
2. **preemptive mitigation:** happens beforehand to avoid the decision event

For example, a "decision appeal process" is a human-initiated, reactive mitigation mechanism. In contrast, a system-initiated, preemptive mitigation mechanism is a transparent explanation for informed guidance for decision-makers before accepting an AI-recommended decision.

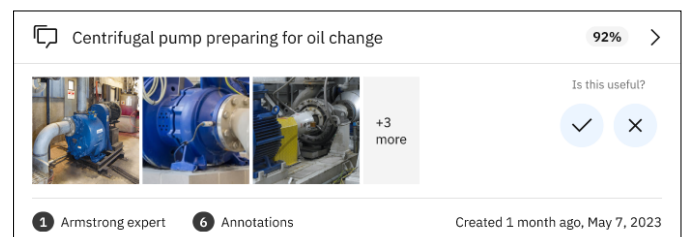
Feedback on output

Enable users to indicate when they disagree with the AI's recommendations or suggestions. Then seek to understand why they differ. This context is necessary to determine what improvements must be made to the model — design ways to seek this context from your users.

Misinterpreted patterns can lead to hallucinations and/or biased output. Therefore it's essential to design mechanisms for AI makers, users, and affected people to point out and correct such misinterpretations. These feedback mechanisms could enable them to point out situational misunderstandings or otherwise indicate deeper disagreement with fundamental concepts learned by your AI system.

For example, an open-ended input response asking why the user disagrees can allow the user to steer the content and volume of feedback but requires higher user motivation to provide feedback. On the other hand, a predefined list of reasons to choose from makes providing feedback a more effortless and quicker action but restricts the context the system receives.

Select a suitable method by considering the severity of the mistake and the amount of context required for its proper fix. Ideally, you want to know whether the user disagrees with the model's learning pattern or a specific situational application of it.



SIMPLE FEEDBACK ON OUTPUT EXAMPLE

Feedback on learning patterns

What might lead to this sort of incorrect learning in AI? To understand this, let's work with an analogy.

Have you ever thought about how children learn social behavior? Sometimes they have explicit guidance in the form of social norms or rules of engagement. But that's not the only way they learn. Often, children also implicitly observe how others react to certain situations and mimic that behavior. In these times, they form their own assumptions about how to respond to similar situations. But, these assumptions can be wrong, as their mistakes sometimes reveal.

Similarly, AI models also learn from explicitly trained as well as implicitly assumed contextual patterns. A case of situational misunderstanding typically means that your model has rough edges that need refining. On the other hand, an incorrectly trained fundamental concept means that your model must be re-wired to address the mistake.

Design valuable and actionable explanations for all users to understand what concepts the AI model(s) applied to arrive at a suggestion or prediction. Because unless they build an appropriate mental model of how the AI system arrived at an output, they will not be able to comment on whether it misinterpreted the context or made a wrong assumption.

Situational misunderstandings might require you to dig deeper into why that situation was special or different. Whereas, with a fundamentally inaccurate concept, you might need to follow up with a dedicated feedback flow to re-train the AI model on the concept.

Balance necessity with choice

Make sure to strike the right tone with your feedback process and ask for user input when it's relevant to the situation. People may not always feel inclined to give feedback unless it's their main responsibility, so focus on building trust during key feedback moments.

Also, be aware of how inaction might affect your AI feedback loop. For example, how will it affect your feedback loop if a user ignores a feedback action you designed? Avoid the slippery slope of incorrectly assuming that inaction implies that a user consistently likes or dislikes the output. Design intentionally for moments when users aren't interacting sufficiently with feedback actions.

If necessary, include user workflows focused on gathering feedback, where you explicitly ask users to review a small set of example outputs.

The goal of designing with a human in the loop, especially through its focus on explainable outcomes and areas to seek feedback from user, is to ensure an intentionally placed set of checks and balances across the lifecycle of an AI system, in a way that it can adapt to user needs and can be appropriately course-corrected with human oversight. By giving it due consideration before hand, and by taking the time to think through the full lifecycle of an AI system, we will lead to more intentional choices every step of the way

Informative insights from data

AI can be a valuable tool to enable human insight – either through providing more information or by helping connect the dots across information. For example:

1. It can analyze large amounts of data to identify patterns and trends that might not be obvious to the human eye. This can be useful in areas such as finance, where AI can analyze market data and help investors make informed decisions about where to invest their money.
2. It can be used to optimize decision-making processes by analyzing multiple scenarios and showing a probability of the most likely scenario to succeed based on given dimensions.
3. It can assist decision-making by providing recommendations based on past data or behavior. (e.g., recommending the most efficient route for a delivery truck based on traffic patterns).
4. It can speed up decision-making by automating context-gathering tasks to quickly fill in the knowledge necessary to make an informed decision.

However, it's important to carefully design this experience with AI to manage potential biases, ensure ethical and fair use, and provide valuable insights that further human information parsing ability. Consider the following areas to guide the design of good information exploration experiences.

Good information parsing

In classical computing, designers help users find and parse information through navigation. Most significantly, good navigation allows users to access content in the most direct way possible. Although there are differences in how users find information with AI, these design principles still apply. Designers need to make it as easy as possible for users to find the right content while allowing them to trace their steps and backtrack regardless of the interface used to present the AI's output.

Humans work with data in two ways – searching to find specific information or broadly exploring data to analyze or discover possibilities. Designers should consider the users' intent when designing information flows in their AI experiences:

- Is the user looking to browse or explore information on a topic? **Focus on communicating** how various data points are related to the topic.

- Or, if they are **looking for specific information**, focus on reducing the time or effort it takes to get the most relevant information available to them.

Provide context for information synthesis

You can't make a good decision without the proper context of the data. In some situations, we look for a complete picture before we make a decision because losing our sense of context can lead us to biased decisions. In other cases, we look for enough of a picture to emerge so we can fill in the gaps. Similarly, when an AI system is used to inform decision-making, understand if the situation and the users involved would want to optimize for completeness of data or if they care more for accuracy or relevance of data.

It's important to clearly communicate this data synthesis context to the user when your AI system is assisting in decision-making through suggestions or recommendations. Without clarification of how an AI arrived at a particular suggestion, it increases user effort to trust the system. However, be sure to understand the most important context to be conveyed so that you're not drowning the user in detail they don't want or need.

For example, imagine a legal or medical case analysis – the practitioners might seek to ensure that relevant data points have been gathered before making a hypothesis. Picture an AI system designed to suggest possible analysis patterns. In this case, it is important to communicate whether or not all possible data points were gathered and analyzed before making a recommendation. Similarly, if the users want, they should be able to see more detail related to any data point as desired.

In contrast, imagine a different scenario in which a support technician troubleshooting a problem with a device is seeking help from an AI system on possible troubleshooting approaches. This user would value a quick response that's relevant to their situation. If multiple responses exist, they want the one that's most likely to succeed. To this user, the exact details of where the answer came from is secondary, so long as it is relevant to their situation.

Design for resilience

We must constantly remind ourselves that nothing is invented and perfected at the same time.

As a designer, anticipating where and how a system or user could fail is essential to designing an experience that can prevent unintended consequences. For instance, issues arising from biases or errors in the data used to train the AI system can be mitigated using a design-for-resilience approach.

By anticipating and planning for potential failures, designers can better understand the limitations and weaknesses of the AI system and take steps to mitigate them before the system is deployed in the real world. Alternatively, analyzing the reasons for failure can help designers identify areas where the system’s training data may be lacking or its decision-making processes may be flawed. This can improve the system’s accuracy, fairness, and transparency.

Fail with intent to improve

Fail·ure /ˈfālyər/ (Noun)

- 1. lack of success.
- 2. the omission of expected or required action.

Put simply, if the user doesn’t achieve their goal or the AI system doesn’t perform an expected or required task, it’s considered a failure.

Many factors could lead to failure, some of which are outlined here.

- **System limitation:** The system can’t provide the proper response or any response at all as it doesn’t know how to deal with this unexpected situation.
- **Contextual errors:** The system is working as intended, but it doesn’t lead to success for the user due to mismatched and possibly incorrect expectations or assumptions on how it was designed to function.
- **Missed opportunities:** Situations where the system behavior was lacking but neither the system nor the human registered an error in the moment.

Given the probabilistic nature of AI systems, it’s even more critical to have a plan in place for when things take an

unexpected turn. We must intentionally design for fault tolerance and graceful functionality degradation when necessary.

Reduce confusion areas for AI

Identifying weak spots or areas where your AI system gets confused and leads to a wrong answer or prediction is a vital activity when designing for resilience. The aim is to train your AI to better identify signals from noise in these situations, thereby reducing the chances of confusion.

For example, consider an AI system designed to detect a certain type of cancer. In most cases, the AI should detect cancer when it’s present (i.e., true positive) or when it detects no cancer when it’s absent (i.e., true negative). In this case, it’s functioning as desired.

However, the AI is “confused” if it doesn’t detect cancer when it *is* present (false negative) or when it detects cancer when there is none (false positive). It’s these confusing situations that we must improve on.

For confusion issues, provide dedicated training cycles initiated by the system or by a human user, where the AI can gain more inputs supervised by a human user and learn to identify the signal from noise correctly. However, remember that not all confusion leads to the same consequences or impact on outcomes.

	AI says Cancer Present	AI says Cancer Not Present
Cancer Present	Hit (true positive)	Miss (false negative)
Cancer Not Present	False Alarm (false positive)	Correct Rejection (true negative)

Some use cases may have high stakes associated with false positives, while others might have high stakes for false negatives. For different situations, false positives and negatives might increase risk. Tailoring your training or feedback plan to your use case leads you to better outcomes without additional unnecessary effort.

When in doubt, fall back to human intervention

To err is human. Similarly, an AI can be characterized by moments of failure. No matter how extensively we train a machine, there's always an edge case, and it won't know what to do. It's prudent to build resiliency by incorporating feedback at such points of doubt. Provide opportunities for a human to assess the situation and determine a course of action.

The goal isn't to eliminate errors but to reduce errors on the part of the users. Our goal is to anticipate how people will respond when things go wrong. Instead of designing an AI system that relies on human-like intelligence — a human-like understanding of language — it's better to design for errors and misunderstandings, hence the term *Designing for Resilience*.

Conclusion

How we design excellent user experiences for AI requires us to expand our thinking. AI requires more and deeper thought on our part to deliver a good user experience that provides purpose, value, and trust to all users.

While we hope to guide designers and cross-functional teams on how to confidently approach their work, this is an evolving practice, just as AI is an evolving field.

Evolving these principles into hands-on practices that guide in implementing these principles into all of our AI experiences towards best outcomes will continue to be an ongoing effort, and we need your input and ideas.

Therefore, we're asking for your help to make this the definitive UX for AI guide for all IBM designers. We welcome your feedback and collaboration to improve upon this body of work. Don't hesitate to get in touch with Emily DiCesaro or Adi Veerubhotla via Slack, and join our Slack channel [#d4ai-ux-for-ai-pub](#) to contribute to this effort.

Appendices

About the team

This document is a product of the [#ai-design-guild](#) at IBM, with contributions from team members across IBM Software. The content in this document was developed by [Dawn Ahukanna](#), [Katrina Alcorn](#), [Jillian Byra](#), [Adam Cutler](#), [Emily DiCesaro](#), [Stefanie Lauria](#), [Mayan Murray](#), [Milena Pribić](#), [Jason Telner](#), and [Adi Veerubhotla](#).

Learn more

Basics

On Bias and Noise

- [Noise; a flaw in human judgment; Daniel Kahneman](#)
- [Thinking Fast and Slow; Daniel Kahneman](#)

AI Bias Amplification

- [This is how AI bias really happens—and why it's so hard to fix](#)
- [Artificial Intelligence May Amplify Bias, But Also Can Help Eliminate It](#)

AI's Propensity to Intensify Current Patterns:

- [Wikipedia entry for Microsoft Tay](#)
- [Meta takes new AI system offline because Twitter users are mean](#)

AI Ethics:

- <https://www.ibm.com/design/ai/ethics/>

Responsibilities of an AI designer

- [AI Quality Heuristics](#)

Informative insights from data

- [Real world examples of confusion due to misleading information context](#)

Designing for Human Perception

- Explainability: <https://www.ibm.com/design/ai/ethics/explainability>
- Interpretability: <https://towardsdatascience.com/interperable-vs-explainable-machine-learning-1fa525e12f48>
- Trust: <https://scienceexchange.caltech.edu/topics/artificial-intelligence-research/trustworthy-ai>

Glossary

A

AI, Artificial Intelligence: The emulation of natural intelligence by a machine. See also: <https://www.ibm.com/topics/artificial-intelligence>

Agency: The degree to which an AI is acting on behalf of its user

AI Infused Applications:

Algorithm: A sequence of unambiguous instructions used by computers to solve problems.

Automation: Automation is the use of technology to perform tasks with where human input is minimized - <https://www.ibm.com/topics/automation>

B

Bias: Systematic error in an AI system that has been designed, intentionally or not, in a way that may generate unfair decisions. Bias can be present both in the algorithm of the AI system and in the data used to train and test it. AI bias can emerge in an AI system as a result of cultural expectations; technical limitations; or unanticipated deployment contexts.

C

Chatbot: A chatbot is a computer program that uses artificial intelligence (AI) and natural language processing (NLP) to understand customer questions and automate responses to them, simulating human conversation - <https://www.ibm.com/topics/chatbots>. Also see Voice Assistant.

D

Data: A recorded snapshot of anything that has happened or is happening. Facts and statistics collected together for reference or analysis.

Data scientist: Combine math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning - <https://www.ibm.com/topics/data-science>

Decision-support processes: An AI system providing recommendations to support human decision-making. <https://www.ibm.com/topics/artificial-intelligence>

Deterministic: When a model's output is completely determined by its inputs and parameter values. See also: [Foundation Models for Designers](#)

E

Explainability: The ability of an AI system to provide insights that humans can use to understand the causes of the system's predictions. See AI Ethics Glossary³ - <https://w3.ibm.com/w3publisher/ai-ethics/knowledge-base/glossary#E>

F

Feedback: The process of providing information or data to an AI system, typically based on the system's outputs or behavior, in order to improve its performance or behavior over time.

Feedback mechanism: An interaction that allows the collection, analysis, and incorporation of information or input from users, stakeholders, or the environment to refine and improve the performance or behavior of an AI system over time.

G

Generative AI: Generative AI refers to deep-learning models that can generate text, images, and other content based on the data they were trained on - <https://research.ibm.com/blog/what-is-generative-AI>

Generative variability: Generative variability refers to the fact that generative models produce artifacts as output and those outputs may vary, even when the input prompt is identical. See [Foundation models for Designers](#)²

H

Hallucination: A situation where an AI system generates or produces information, data, or outputs that do not correspond to reality or exhibit characteristics that are not present in the input or training data. It can involve the AI system creating imaginary or fabricated content that appears authentic or generating inaccurate and misleading results.

Human-in-the-loop (HitL): Human-in-the-loop machine learning has two distinct goals: making a machine learning application more accurate with human input and improving a human task with the aid of machine learning. we're focusing on human task improvement with the aid of machine learning - <https://livebook.manning.com/book/human-in-the-loop-machine-learning>

I

Intelligent recommendation system: Recommendation systems are a subclass of information filtering systems that use AI to predict the rating or preference a user would give to an item.

L

LLM, large language model: A type of AI algorithm that uses deep learning techniques and huge data sets to understand, summarize, generate and predict new content. See also: [Foundation Models for Designers](#)

M

Machine Learning (ML): A subset of AI that enables machines to develop problem-solving mathematical models by identifying patterns in data instead of leveraging explicit programming.

Mental model: An individual's understanding of how a system works and how their actions affect system outcomes. These expectations often do not match the actual capabilities of a system which may lead to frustration, abandonment or misuse. See [Foundation models for Designers²](#)

Model: A function that takes features as input and predicts labels as output.

P

Predictable: Able to be known, seen or declared in advance.

Probabilistic: The characteristic of being subject to randomness; non-deterministic. Probabilistic models do not produce the same outputs given the same inputs. See also [generative variability](#).

R

Reliance: The degree of trust or dependence placed on AI systems or algorithms to make decisions, provide information, or perform tasks. It indicates the extent to which individuals or organizations rely on AI technologies for critical functions, taking into account their accuracy, consistency, and overall performance.

Resilience: The ability of an AI system or algorithm to adapt, recover, or continue functioning effectively in the face of challenges, disruptions, or unexpected circumstances.

T

Transparency: Users must be able to see how the service works, evaluate its functionality, and comprehend its strengths and limitations. See <https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas> and <https://w3.ibm.com/w3publisher/ai-ethics/knowledge-base/glossary>

Trust: Trust in AI relies on understanding how it works. Users and Designers need to understand why AI makes the decisions it does. AI that helps to build trust through a human-centric approach characterized by five requirements: explainability, fairness, robustness, transparency, and privacy. See <https://research.ibm.com/topics/trustworthy-ai> and <https://w3.ibm.com/w3publisher/ai-ethics/knowledge-base/glossary>

U

User Experience for AI (UX4AI): The overall experience of a person using an AI-based digital product such as a website or computer application, especially in terms of how ethical and easy it is to use.

V

Voice assistant: Voice control, also called voice assistance, is a AI conversational user interface that enables hands-free operation and task completion with a digital device. Also see [Chatbot](#).

Resources and links

Templates: General

- [EDT for Data and AI template](#)
- [EDT for Conversation Design template](#)
- [AI Competitive Analysis template](#)
- [AI Quality Heuristics template](#)

Templates: Ethics

- [EDT for AI Ethics: Core Template](#)
- [EDT for AI Ethics Question-based Explainability Workshop template](#)
- [EDT for AI Ethics: Ethical Pillars template](#)
- [EDT for AI Ethics: Future Narratives](#)
- [Ethics by Design AI FactSheet template](#)
- [EDT for AI Ethics: Explainability Mental Model Mapping template](#)
- [EDT for AI Ethics: Ethics Canvas](#)

Sites

- [IBM Design for AI](#)
- [Design for AI Guild Plenary Meeting sign up](#)
- [Design for AI Guild Speaker Series sign up](#)
- [IBM Sustainability Software AI Design site](#)
- [IBM AI Ethics site](#)
- [Everyday Ethics for Artificial Intelligence](#)
- [IBM Natural Conversation Framework](#)
- [Conversation Design Best Practices](#)
- [Data & AI Design Framework](#)
- [Library of Applied Explainability Airtable](#)
- [UX for AI Principles - draft version](#)

Primers

- [Foundation Models for Designers](#)