dscout

The New Properties of GenAI

(and How to Study Them)

GenAI is tricky, probabilistic consequential, and on top of all that—constantly evolving. Learn how you can stay up-to-date and keep the user at the forefront of product development.

Generative AI has arrived and whether you like it or not, it's here to stay.

At dscout, we've been thinking a lot about Generative AI these past few months. We've interviewed experts on how it's changing the UXR role, hosted panels on the benefits and risks, and thought deeply about the impact that research will (or should) have on this newest tech boom.

In all of these conversations, one thing is clear: **user research must play a critical role in the development of Generative AI products.**

OK, but *how*?

You may be running into this question in your work. Perhaps your organizations have given you some broad directives, like "start researching AI," but don't know exactly what they're looking for.

Perhaps some of your companies are not researching as much as they should, or they invited you to the process too late and you're sprinting to catch up. You want to advocate for UXRs in GenAI development, but you don't have a clear vision of what exactly the research agenda should be.

Situations like these are very understandable. After all, Generative AI isn't just any new technology. It has novel, paradigm-shifting elements that are changing the way that builders build and the way users interact with technology. Research needs to adapt as well, but with all these new unknowns, it's difficult to know where to start.

We wanted to orient ourselves in this new landscape. So, we spoke to [11 seasoned AI researchers](#) about what questions they think are important to ask about Generative AI tools, and what methods they've been using to answer them.

We brought some of these findings to [dscout's Co-Lab earlier this year](#), where we continued the conversation about Generative AI with research leaders from around the country.

Through all these discussions, we learned that there are [four novel properties](#) of Generative AI tools that UXRs are talking about. Combined, they make GenAI very different from other tools that UXRs typically study.

1. It's an **unknown variable**—we don't fully understand its function or its boundaries, and we can't explain them to our users.

2. It's **probabilistic** (non-deterministic)—we cannot guarantee any particular outcome. Errors aren't edge cases anymore.

3. It's **consequential**—human emotion and relationships are entangled, and we don't fully understand how, or what the implications will be.

4. It's **nascent**—users have very little frame of reference for GenAI tech. Their mental models are changing every day.

In this article, we'll talk about these four properties and how experts are starting to learn more about each one. We also will hand you some ideas to build out your own research agenda for 2024.

GenAI is an unknown and unknowable variable

The first unique property of Generative AI is fascinating and a little troubling: the exact mechanism of how Generative AI makes predictions is extremely challenging to understand, let alone control.

Of course, we know how GenAI algorithms work on a broad scale—it's a lot to go into here, but you can read some fantastic primers on how models make predictions.

However, the actual variables that are being shifted as GenAI models "think" are more or less hidden, even to their builders. They are also apt to change as the algorithm continues to train on more data. This makes understanding how a model arrives at any individual output next to impossible.

For example in 2022, researchers studied GPT-2 (a significantly simpler model than our current ones) and published a 25-page paper on how it predicted one word at the end of one sentence. So yes, a prediction was technically traceable. But at the end of the day, that process is not scalable at all, functionally rendering the inner workings of algorithms like GPT-4 unknowable.

Then layer on the fact that, with few exceptions, organizations are not building and using their own models. This new technology suddenly became available to everyone at once and we're all rushing to experiment with it and figure out how it works and what it can do.

Our organizations don't have visibility into how the model was trained or what changes are being made by its builders, both of which can drastically change the function of our end tools.

This presents an interesting new paradigm for organizations. It feels weird to even have to articulate this, but with non-AI-powered tools, we know what we are giving our users. After all, we're the ones who built it. And if something goes wrong, we know what to tweak to fix it. We have the ability—and responsibility—to be accountable for the output of our tools.

But when it comes to AI, and Generative AI in particular, we can't open up the hatch and tinker with our tools in the same way. We can't even fully understand or predict how it will behave to the end user. **But the accountability to provide useful, reliable, and safe tools remains the same**.

Researching the "unknown": Responsible use cases

UXRs can play several important strategic roles in helping design around this unknown and unknowable property of GenAI.

We can build expertise on the affordance of our tools

Our builders also just got handed this tech, and they could probably use extra help in experimenting with what's possible. What can and can't be done with the GenAI tools at your disposal? What can be done reliably? How can prompts be tweaked to make it more reliable?

Building these kinds of expertise will take some doing since most of us don't have a very technical background. However, building that expertise could have a long-term payoff for your research function. Working on the more technical side might also provide an opportunity to build relationships with your engineering teams.

We can keep our organizations accountable to the unknown

Learning how to build with GenAI tools is exciting and the end capabilities can be dazzling. As teams discover more about how the tools might function in the context of their products, they might accidentally put on some rose-colored glasses.

As the voice of the user, we can both encourage the exploration of these tools while holding our organizations accountable to the core truth: we can't fully predict or guarantee an end-user experience. We don't know when GenAI tools might show bias or pose risks we hadn't previously considered.

This doesn't necessarily take much novel research. It can be as simple as being the voice in the room that reminds others of the slippery, unknown nature of these tools. We can ask our organizations the more difficult questions: are we sure this new GenAI tool will be a benefit and not a risk to our users? Are we willing to gamble on it when we can't guarantee how the tool will function? And are we willing to accept the consequences if it behaves in a way we didn't predict?

> *Liz Jernegan (Senior UX Researcher at Amazon Web Services) spoke in depth about keeping her teams accountable to the unknown in her talk at Co-Lab this year, [watch her presentation here](#).*

We can explore use cases that are effective and responsible

We don't just have to be asking tough questions. We can give great answers, too.

UXRs can play an important role in determining what GenAI should do. One of our main goals should be to gauge what we can confidently roll out, knowing that we cannot fully guarantee or explain the experience that users will have with whatever we build.

We can do so by establishing new use cases for GenAI that meet known needs in our user base while minimizing risk. As one of our experts put it, we can make sure that what we are delivering is not just different from our old tool, but different and better.

Establishing use cases quickly is one way to get the UX voice into the product development conversation. If we're not there, the "problems" that end up getting solved by GenAI might be created in brainstorms by product teams who are well-versed in AI's potential, but not familiar with the standard users' experiences. This can end up with "solutions" that don't actually resonate with users at all.

We can avoid this if we hand them data-driven use cases that spark responsible and efficient innovation. But to do this, we need to have up-to-date GenAI use cases to bring to the table, before devs get to work brainstorming their own.

[The seasoned AI researchers we spoke to](#) are meeting this goal by returning to basic foundational research practices. This included revisiting well-trodden use cases to see what might be useful for users without being potentially dangerous or difficult to control. It also included running new exploratory research to identify potential new personas and use cases.

Researching the unknowable: What can you do?

- **Explore the tools available:** Get into any GenAI tools you have access to and start playing around with them. Understand their capabilities and try to develop an intuition about what they're good at.

- **Keep learning:** Read up on how GenAI works and what its limitations are.

- **Research GenAI bias:** See for yourself how it's created. Be curious (and skeptical) about how your organization's models are being trained.

- **Revisit research:** Explore old use cases, personas, or other foundational research, to see if there are gaps in your functionality that might be newly feasible with GenAI.

- **Foundational and exploratory research:** Conduct new foundational and exploratory research to generate GenAI use cases for your company—try Jobs to be Done, for example.

- **Be realistic:** Assess what you can reliably and responsibly deliver to your users.

- **Assess the risks:** Encourage your teams to think thoroughly about the risks associated with any new tools. Try to imagine what the "worst case" would be for a use case gone awry, and not just the best-case outcomes. If running foundational research, ask your users about potential risks and fears, and not just potential benefits.

GenAI is probabilistic

Up until now, the vast majority of technology put to market has been **deterministic**. That is, when users input X, tools will always output Y. But GenAI works in a **probabilistic** model, which is better summed up as "If X, then most likely Y."

This is an inherent property of Generative AI tools. To simplify it way down, Large Language Models are extremely advanced text prediction machines. That means that choosing the next word or sentence is always going to come down to an educated guess. And LLMs (however smart they appear) have no common sense to gut-check their answers. This is why we sometimes get "hallucinations" or other nonsense output from a normal prompt—on some level, it's just a roll of the dice.

> *Learn more about how text prediction works in Kevin Johnson's (Head of AI at dscout)* [Co-Lab Continued session](#)*.*

This is an interesting shift for users, who are used to almost exclusively deterministic output in their products. This is bound to cause friction as users start to interact with GenAI tools. Usually, when something doesn't give us what we want, we assume it's broken. But that's not the case with GenAI, which is acting as expected even when it produces errors.

There's the added issue that, if users don't understand that something isn't deterministic, they may put more trust in a GenAI output than it deserves. They might not double-check an answer, for example, or proofread an auto-generated email before sending it somewhere. And if that ends up being a mistake, our organizations will likely bear the weight of that frustration.

This is a paradigm shift for designers and researchers as much as it is for users. "Errors" are no longer bugs or edge cases—they are a core element of our tools. We must embrace a new philosophy of product design that fully incorporates error and hallucination as part of the expected output.

Researching probabilistic GenAI: Building and maintaining trust

UXRs have thorny questions on our hands with probabilistic design. How do we encourage users to engage with GenAI tools, while also telling them that they can't always rely on the output? How do we build and maintain positive sentiment in the face of errors? And how does probabilistic output maintain brand trust overall?

Our seasoned AI researchers are addressing these questions tactically through **feedback and metrics**. They are keeping fresh tabs on how users are interacting with GenAI, and how their interactions affect brand trust overall. And they're taking that feedback into consideration when designing communication around probabilistic output.

When it comes to generative AI, **the traditional metrics measuring product success are as good as useless.** Researchers are returning to the drawing board to figure out what measurable success looks like in a product that's so qualitative and contextually dependent. Others are starting to consider building new feedback loops into their system to try and gather more meaningful feedback on Generative AI interaction in their product.

Researching probabilistic GenAI: What can you do?

- **Partner with engineering:** Work closely with engineering and QA to understand what errors and hallucinations look like in your base tool. What does bad output look like, and how frequent is it?

- **Define success:** Think about what "success" means for your output, and what kind of feedback you need to track it. Establish new metrics and feedback loops in your product to start tracking it.

- **Run usability tests and surveys:** We can expand the notion of "usability" to include more than usefulness to helpfulness, safety, etc. We can use these methods to determine how much users trust the output they are given. Compare the actual error rate to understand how users' mental models compare to your actual product.

- **Field test:** Test different concepts for communicating the probabilistic nature of your tool to your users.

- **Run evaluative research:** Use this to determine how much bad output impacts brand trust overall.

GenAI is socially consequential

This is a big one and maybe the most important new property that we need to be paying attention to.

Generative AI can't just be treated like "only technology." LLMs are very successfully able to mimic human speech and interaction. In short, it feels 'human.'

It's important to note that this does not mean GenAI is sentient. It is, at its core, still a complicated text prediction machine. But the mastery of human-like language use is hardwired to resonate with us and sets us up to assume a sense of personhood and sentience when none is there.

When we believe something to be human (even on a subconscious level), we begin to trust it, form relationships, and potentially be swayed much more easily by its suggestion. This leaves consumers of Generative AI content socially vulnerable in ways we haven't seen before with less "human" tools.

Users might be more inclined to trust its output, share things they wouldn't normally be comfortable sharing, or even get emotionally attached to your tool. This is especially true if your company has tools with a friendly or relatable brand personality, who might be looking to inject that into their AI tools.

We are blurring the lines between human and machine interaction, and we know very little about what the implications of that are. But from what we have seen so far, it should be clear that we are playing with fire, and we must proceed with as much information as possible.

Unfortunately, there aren't many other roles in the organization that are tasked with thinking about the social implications of GenAI tools. Many organizations might be thinking that the best possible design choice would be to make their tool as convincingly human as possible and to make the transition between human and machine feel invisible. This is in line with design principles of making a 'seamless, delightful' experience. But with this new socially consequential tool, may end up being dangerous.

> *Learn more about social consequences of AI in [Katie Johnson's (Fractional Head of Research at Yohana) Co-Lab Continued](#) session.*

Researching GenAI: Relationships, emotions, and imagination

UXRs have often been put in the awkward position of being the "Voice of the User" or the "Empathy Engine" when it comes to building new products. This tension is highlighted here in the realm of social consequence in generative AI.

We have the opportunity to shape these tools in a way that's emotionally safe for our users, but to advocate we need to know a lot more about how emotional relationships and bonds between user and AI develop. We need to understand how people are emotionally contextualizing and interpreting interactions with GenAI, and how those emotions are influencing user behavior.

Social consequence also brings up important design questions around transparency. Historically, seamlessness has been the name of the game when it comes to design. But is that still the right thing to do? When many LLMs can successfully imitate human interaction, how do we keep the line crisp between human and machine? What are the consequences if we don't?

Questions like these are important to protect our users, but they will also have business implications. If LLMs end up crossing uncomfortable boundaries with the user, that could seriously impact how they feel about the tool as a whole. Nobody wants to interact with a creepy uncanny valley robot, after all.

How do we keep our GenAI feeling useful, perhaps friendly even, without crossing the line into creepy or invasive? To answer these questions, we need to get specific about what kinds of interactions influence user sentiment, so that we can guardrail around them.

Of course, as we all know, studying emotions and relationships is hard. Questions like "When did this start feeling like a person to you?" are abstract and hard to answer. Answering specific questions about what felt good, bad, or creepy in a past GenAI interaction will be difficult to recall, especially if the interaction was in the past. We need to get more specific and actionable with our findings if we are going to design against them.

There's also the issue of sensitivity. People might be sharing information with GenAI that they don't necessarily feel comfortable sharing with a researcher. Or they might be embarrassed or unwilling to share their emotions around their interactions with GenAI, especially if it's about something already emotionally vulnerable, like finances, health care, or personal struggles.

Researchers who study GenAI are actively grappling with how to tackle this challenge. Some have tried diary studies to get direct feedback on individual interactions. For example, at Co-Lab Continued, Meredith McDermott of Duolingo told us how she used a more open-ended qualitative approach to get feedback on her new GenAI product using dscout Diary.

> *See how Duolingo created a unique, fun (and non-awkward) tutoring experience with GenAI in Meredith McDermott's (Staff UXR at Duolingo) session at Co-Lab Continued.*

Others are returning to immersive ethnography and participant observation. One researcher we spoke to is trying biometric data to get real-time stress response feedback during interactions with GenAI in a lab.

We're in the wild west when it comes to researching human-AI social interaction. We're going to have to get creative in the coming months and years with just how we approach these challenging relational questions.

Researching social consequences: What can you do?

- **Research relationships:** Take a step back from your organization's products and do some background reading on how humans form relationships and build trust, whether with other humans or with human-like language models.

- **Diary studies:** Run diary studies of users interacting with GenAI tools. Pay attention to how they're conceptualizing the tool, what kind of language they use when talking to and about it, and how it affects their level of trust and engagement with tool output.

- **Identify acceptable vs. creepy:** Run research to understand what, specifically, your product does to seem 'sentient' or not. What is moving the needle on people's mental models of your tool? What kinds of machine behavior feel acceptable to users, and when does it get creepy? Try to isolate the concrete variables that contribute to the uncanny valley.

- **Research what makes something human-like:** Test hypotheses about what makes something feel 'human' with comparative concept tests. Change the language and framing around your tool and see how sentiment and mental models shift.

- **Advocate for clarity:** Push product and design efforts to delineate when people are talking to a machine versus a real human.

GenAI is constantly evolving

The last novel element of GenAI that we need to pay attention to: It's just so new.

But it affects our users, too. The public has very little frame of reference for how to relate to this new technology. The researchers we spoke to are predicting that these mental models are likely to change quickly. This will probably happen both on an individual level, based on personal interactions with GenAI tools, and on a societal level as the discourse about AI continues to evolve.

Users' lack of a stable mental model can cause misunderstandings about what GenAI is and what it can do. This compounds the usability and ethical concerns related to the other novel elements of GenAI:

- Not understanding how unknown the inner workings are leads to treating LLM output like a source of truth, not recognizing bias, and giving up more data than intended.

- Not grasping probabilistic output means trusting output without double checking and not being aware of hallucination issues.

- Not understanding the social consequences leads to users assuming LLM sentience or humanity, confusing machine and human interaction, and forming unhealthy social bonds.

Researching evolving AI: Awareness and education

UXRs will have their work cut out for them keeping up with the constantly evolving technology and the public's changing understanding of it. Documenting how users are imagining these tools, and how that's shifting over time, will be crucial information for organizations as they pivot to stay on top of trends. We have a key role to play in this, but it will mean staying nimble and keeping our finger on the pulse of public conversation.

Longitudinal studies: Our seasoned researchers are responding to this need by increasing the amount of longitudinal studies. Methods like repeating benchmarking surveys or long-form diary studies are helping them to see how user understandings of GenAI are evolving in real-time.

Passive data collection: This tactic is also becoming a bigger part of some researchers' landscapes. They are making closer relationships with their data science and engineering teams, and working together to make sense of evolving metrics in GenAI tools.

Education: Lastly, researchers are looking at the related issue of education. Given the current perception and awareness of LLM functionality, what do users need to understand about the tools we're building to use them safely and effectively?

Researching an ever-changing technology: What can you do?

- **Stay focused on the user:** Never take your finger off the pulse of user understanding! Start early adopters in long-term studies or run quick-hit repeating express missions to understand knowledge levels and attitudes around AI.

- **Keep your team updated:** Bring insights about user perceptions to your teams as regularly as you're collecting them. Keep your product teams accountable to the average users' amount of knowledge about GenAI.

- **Stay in the know:** Keep on top of the news and discourse about GenAI to understand larger societal perceptions of new technology.

- **Do exploratory research:** Understand what users need to know to interact with your software responsibly.

- **Start workshops:** Engage in internal workshops to design educational materials about your new GenAI products.

GenAI is moving quickly, and as user researchers, we have a great opportunity to be at the forefront of this technology and use it to benefit the user, our organization, and our practice. To read more about why GenAI needs UXR and vice versa, take a look at Michael Winnick's (CEO at dscout) article [GPT Needs UXR](#).

———————————————

See how dscout is developing AI tools that work with you (not in place of you)

If you're interested in learning more about how dscout is approaching GenAI research and the tools we're building to support your work, see a preview of what we're working on.

Read all about it