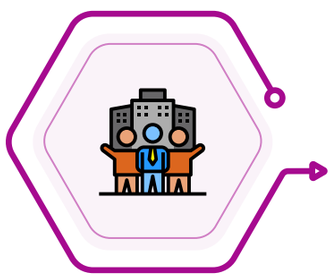
The Rulebook for Designing AI Experiences

Ask ten people in tech what responsible AI design means and you will get ten different answers. Ethics frameworks, trustworthy AI principles, responsible innovation checklists — the vocabulary keeps expanding. But beneath all of that lies a simpler question: when someone is sitting in front of an AI-powered product, what actually makes the experience good?



It turns out three organizations have tried to answer that. Over the past several years, Microsoft, Google, and IBM have each published frameworks for Human-AI Interaction, or HAI, design. These are not just high-level ethics statements but concrete, pattern-based resources for teams shipping real products. The three companies do not always agree, and none of them covers everything, but together they form a body of work worth understanding. What sets them apart is their attempt to bridge the gap between principle and practice — a gap that is wider than it looks.

Microsoft: 18 Guidelines and a Design Library

Microsoft's contribution is the HAX Toolkit, built around 18 guidelines for Human-AI Interaction. First introduced in a 2019 paper by Saleema Amershi and colleagues at Microsoft Research, the work has become one of the most cited references in the field.

01

The HAX Design Library is where these principles come to life. Each guideline is paired with design patterns and real product examples, filterable by product category, AI type, and design goal. Goals include transparency, personalization, reliability, fairness, and appropriate reliance, so teams can quickly find what fits their context. The guidelines span the full lifecycle of an interaction: setting expectations early, recovering gracefully from errors, and managing how a system learns over time and notifies users when its capabilities change.

02

One of the toolkit's strongest ideas is appropriate reliance. The goal is not simply to build trust — it is to build the right amount of trust, calibrated to what the system can actually do. A medical triage tool that delivers suggestions with the confidence of a music recommender is not just poorly designed; it is potentially dangerous. Appropriate reliance helps users sense when to defer to AI and when to push back.

Google: Starting With the Right Questions

Google's People + AI Guidebook, produced by the PAIR research team, takes a different approach. Instead of numbered guidelines, it is structured around six themes: User Needs and Defining Success, Mental Models, Feedback and Control, Explainability and Trust, Data Collection and Evaluation, and Errors and Graceful Failure. Each chapter includes worksheets that turn guidance into team activities.

01

The 2021 second edition added standalone design patterns organized around the questions teams actually ask: How should I use AI in my product? How do I onboard users to new AI features? How do I explain my AI system? That shift reflects how designers really work — dipping into a resource at a specific moment rather than reading cover to cover.

02

One pattern stands out for conversational interfaces: explain for understanding, not completeness. When surfacing AI reasoning, focus on what users need to move forward rather than exposing every layer of logic underneath. People want enough to feel informed, not a technical lecture mid-task. Traffic to the Guidebook jumped 560 percent between February and August 2023, as generative AI products flooded the market and teams needed something more grounded than philosophy.

IBM: Ethics as Infrastructure

IBM approaches the territory from two directions. Their design practice rests on AI design ethics principles meant as a shared foundation across all IBM products, covering accountability, explainability, value alignment, fairness, and user data rights. The companion AI Essentials Framework, a team-facing tool, is built on five pillars: intent, data, understanding, reasoning, and knowledge.

01

Their more provocative contribution came at CHI 2024, where IBM Research presented six design principles for generative AI. The paper argued that most existing HAI frameworks were built for AI that makes decisions — classifying, ranking, predicting. Generative AI is different. Rather than reaching a conclusion, it produces something: a draft, an image, a block of code. The design challenge shifts from helping users evaluate an output to helping them shape one.

02

Three of the principles revisit familiar concerns through a generative lens: designing responsibly given new risks of misinformation and bias, designing for accurate mental models, and designing for appropriate trust when systems can produce confident-sounding nonsense. The other three address genuinely new characteristics. Designing for Generative Variability accepts that the same prompt can produce different outputs, which runs counter to UX traditions of consistency. Designing for Co-Creation gives users controls to actively steer the generative process. Designing for Imperfection calls for transparency about results that may be plausible but wrong, with mechanisms to flag and fix flaws.

Common Ground and Real Gaps

All three frameworks converge on the same core concerns: transparency about capabilities and limits, support for user control and correction, feedback loops that improve both sides, and graceful failure when things go wrong. The fact that they were developed independently and arrive at similar foundations suggests those foundations are reasonably solid.

The gaps tend to show up at the edges. Research comparing the Microsoft and Google resources against the EU's Ethics Guidelines for Trustworthy AI found that diversity, non-discrimination, fairness, and environmental and social well-being are relatively underserved. Part of this is structural — frameworks built by large tech companies reflect the problems those companies are trying to solve, which often means usability and error recovery more than broader social impact. Fairness and environmental concerns are also harder to translate into design patterns. There is also the generative AI gap IBM identified: frameworks built for an earlier generation of AI do not stretch naturally to systems that write, draw, and generate.

Should You Use These?

The honest answer is that these frameworks work best as thinking tools, not compliance checklists. None will tell you exactly what to do in a specific design decision. What they offer is a shared vocabulary, a set of dimensions to pressure-test against, and the grounding to push back when a product is optimizing for capability at the expense of user comprehension.

In practice, that might mean running a heuristic review against Microsoft's 18 guidelines before shipping, using the PAIR worksheets to structure a critique, or aligning on intent and data strategy with IBM's AI Essentials Framework before design begins. Used lightly and honestly, they work more like a lens than a process. All three are free, well-maintained, and built on serious work — a reasonable place to start, even if not the end of the conversation.