Experts in the Loop: Why Humans Will Not Be Displaced by Machines when there Is "No Right Answer"

TOM HOY, *Stripe Partners*
JANNEKE VAN HOFWEGEN, *Google*

This paper explores the resilience of human expertise in an age of Large Language Models (LLMs) and establishes a framework and methodology for the ongoing value of human experts alongside conversational AI systems. Our framework identifies three themes that define the unique value of human expertise: curating knowledge, personalizing interactions, and offering a perspective. For each theme, we outline strategies and tactics used by human experts that AI struggles to meaningfully replicate at scale. We contrast this with complementary strategies that AI systems are uniquely positioned to offer. Finally, we provide design and engineering principles to guide product teams seeking to shape the implementation of this emerging technology in a way that augments, rather than automates, human experts across consumer domains. The framework and methodology are described in practice through a case study of travel planning—a domain where there is "no right answer".

In the second half of the 20th century, the field of Artificial Intelligence (AI) began developing what became known as "expert systems". Expert systems were hand-coded to automate narrow, repetitive tasks (Dreyfus 1987). These systems are useful when the "right answer" is universal and can be coded from textbooks and experts, such as a system making medical diagnosis for specific conditions, or a system designed to judge whether to approve a loan based on fixed rules (Buchanan et al. 2006; Forsythe 2001).

In the 1980s, new AI architectures structured around neural networks could both develop internal representations of knowledge and learn patterns in data. As computing power increased in the following decades, these innovations meant systems were not only defined by hand-programmed rules, but could operate as "learning systems" enabled by vast amounts of data. Tasks such as fraud detection and weather prediction could now be conducted more autonomously, with less need for humans-in-the-loop.

In recent years, these learning systems began to automate elements of more complex tasks such as spam detection, language translation and facial recognition. But these systems complemented or accelerated the work of humans rather than replaced them, especially tasks that required the expertise of "white collar" knowledge workers.

Human experts remained uniquely suited to tasks that addressed ambiguous, emergent and complex problems that required novel, iterative approaches and original thinking, problems which may have nuanced answers, or no "right answers" at all. As Frey and Osborne summarized back in 2013, AI systems were still limited to relatively narrow tasks that enable "a programmer to write a set of procedures or rules that appropriately direct the technology in each possible contingency. Computers will therefore be relatively productive to human labor when a problem can be specified—in the sense that the criteria for success are quantifiable and can readily be evaluated."

This narrative shifted significantly in October 2022, when ChatGPT was launched. The adaptable, generative intelligence demonstrated by this new technology gave rise to claims that AI systems are, for the first time, capable of displacing human expertise in these more complex and creative domains. Speculation subsequently arose about the threat of automation to jobs previously believed to be immune to AI systems, such as copywriters, journalists, designers, customer-service agents, paralegals, coders, and digital marketers (Herbold et al. 2023; Lowery 2023).

To what extent, though, is human expertise being replaced in reality? Our research into the value of human expertise in one domain, travel discovery and planning, suggests we should be cautious about these broader claims. Travel is frequently used as an example to illustrate how conversational interfaces powered by Large Language Models (LLMs), such as Google's Gemini (Hawkins 2024) and OpenAI's ChatGPT (Mountcastle 2024), are transforming complex tasks and decisions that would have previously relied on human expertise. Travel shares a number of underlying characteristics with other domains that makes it a useful case study for exploring human expertise, and the extent to which it is automatable.

"No Right Answers" as the Contested Space for Displacement

Exploring travel advice reveals deeper truths about human expertise because travel is a domain in which there is "no right answer" (Pierce 2023). For each traveler there are multiple destinations, lodgings and itineraries that can address their desires and unique context. This makes decision making particularly ambiguous. As The Verge puts it "there's no page on the internet titled 'Best vacation in Paris for a family with two kids, one of whom has peanut allergies and the other of whom loves soccer'" (Pierce 2023).

Markets in which there is "no right answer" contain products that Economic Sociologist Lucian Karpik defines as "singularities" (Karpik 2010). Singularities are

unique products defined by three qualities: multidimensionality, uncertainty and incommensurability. Alongside travel other archetypal examples of "singularities" include luxury goods, wine, literature, music, movies, art, dining out, antiques, real estate and professional services.

Travel is a multidimensional product because it has many aspects and layers—from the taste of the local food to the comfort of a hotel room—that a traveler must try to assess. Travel is uncertain because, no matter how many user-generated reviews one reads, one can never fully know what the real experience of a vacation will be like in advance. And travel is incommensurable because no two vacations can be impartially compared: there is no objective basis on which to say a city break is better than a weekend in the mountains. Travel choices are contingent because they are a matter of taste, preference and circumstance.

This has important implications for how people make decisions about their vacation. Because there is "no right answer", the "market of singularities requires knowledge of the product that far exceeds anything necessary for the standard market" (Karpik 2010). To acquire this knowledge, would-be travelers rely on a variety of "judgment devices" to inform their decisions about travel: social networks, trusted brands and accreditations, critics and guides, expert and buyer rankings, and established sales channels (Jacobson and Munar 2012).

Because travel is an unknowable singularity, travelers must trust judgment devices to help them make their decision. This is where, historically, human expertise has been critical (Confente 2015). Alongside advice from friends and family (social networks), travelers rely on the inspiration and recommendations provided by travel agents, travel journalists, bloggers and influencers (critics and guides), as well as user reviews (expert and buyer rankings).

Today, human expertise published in the form of social media posts, user reviews or articles is often found intentionally by travelers via search, or discovered passively via algorithmic feeds. In both cases the results are tailored to the user in question. However, until now no digital product has been able to make this advice truly "personalized". For example, while Instagram may be able to predict the types of vacation a user might be interested in based on their profile and behavior over time, it does not know the full context and requirements of a specific trip, nor can it iterate recommendations via ongoing feedback.

A truly personalized product is, by definition, the most unique type of product possible because it is tailored to one customer (or customers). Karpik defines it as the "pure[est] form of singularity" (Karpik 2010). This is why, historically, many

travelers seek direct, personal advice from other people. To date, only other humans have come close to offering truly personalized recommendations.

For the first time, LLM technology might change this. Today it is possible to ask ChatGPT what the "Best vacation in Paris for a family with two kids, one of whom has peanut allergies and the other of whom loves soccer" is. And ChatGPT will confidently answer. This confidence stems from the large amounts of web data that a language model is trained on to "understand" the user requirements (family-friendly, peanut allergy, soccer) specified in their query, match it against the most relevant sources in its training data and synthesize an answer. In recent models such as Gemini, this black-box process is made traceable by the LLM supporting its generated answer with direct online sources, further adding to the sense of confidence of its prediction. Does this threaten the livelihoods of travel experts who earn a living by supporting travelers addressing these specific requests? And what, by extension, does that mean for human expertise in other "no right answer" domains?

A Multidisciplinary Approach to Understanding Expertise

The Google Search team is exploring the opportunities afforded by generative AI technology, including Google's own Gemini model, to better serve Search users. It was within this context that a cross functional team of Google product directors, designers and engineers tasked a joint Google and Stripe Partners research team with exploring the nature and value of human expertise in travel, as a case study to inform the new Conversational AI and generative search interfaces they were building. By understanding what expertise was uniquely human, Google could define the scope of emerging products, focusing on where a Conversational AI can add complementary value to existing forms of expertise.

The joint Google and Stripe Partners team consisted of qualitative researchers, linguists and data scientists. Together we designed a novel mixed-methods research approach combining ethnographic observation and in-depth interviews with NLP (Natural Language Processing)-driven analysis.

Our first decision was how to focus the study. We selected a single, versatile travel destination, Los Angeles (LA), as the focus of our study because LA is one of the most commonly searched-for and popular travel destinations in the US (Chang 2023), and offers a diversity of common travel-related activities (from museums to night life to beaches).

We recruited 9 "LA travel experts" ranging across three types of expertise. The first category were LA-focused travel agents who earned a living recommending and selling trip packages directly to travelers. We included travel agents in our sample

because we wanted to capture the value of "professional" travel experts who earned money directly from travel recommendation and planning. Second, we recruited social media influencers who frequently shared LA-related recommendations and content across their channels to their audience. We included influencers in our sample because we wanted to capture people who are earning money indirectly from their LA-related expertise. And finally we recruited LA locals; individuals who had lived in LA for at least 5 years, considered themselves local experts, and frequently shared advice informally to friends and family interested in visiting LA. We included locals because we wanted to explore the value of advice not influenced directly or indirectly by financial incentives.

We paired each expert with a tourist intending to travel to LA within 6 months and arranged a 30 minute "planning and advice session" between them via video call. In addition to recording their natural language conversation, we observed body language and tone of voice to identify pivotal interactions. Observing how experts and travelers interact was critical because, as Diana Forsythe argues in her renowned study of AI engineers attempting to acquire knowledge, experts "are not completely aware of everything they know" (2001).

Following the session we conducted an in-depth interview with the traveler to capture immediate reflections. Our qualitative research team then conducted a "grounded theory" style driven analysis of the 9 advice sessions and traveler interviews to surface common patterns across traveler needs expert strategies.

In parallel, our data science team ran NLP analysis of the sessions. This computational approach is complementary to qualitative observation and interviewing because it provides a detailed breakdown of the turn-based conversation between traveler and expert, allowing us to understand more precisely what characteristics of the language (outlined below) lead to the most successful dynamics. The approach is especially relevant here because we are trying to understand the nuances of conversational interaction, with a view to understanding its implications for a Conversational AI interface.

The NLP analysis was conducted across all interviews, resulting in thousands of expert advice-traveler response pairs which are analyzed at scale to quantify emerging patterns in our interviews.
The computational linguistic analysis included:

- Temporal analysis: (a) Spoken duration for expert and traveler (b) Average pause taken by expert/traveler before responding to traveler/expert (c) Intimate turns initiated by expert/traveler (d) Questions asked by expert/traveler

- Sentiment analysis
- Formality analysis
- Intimate sentence identification
- Specificity analysis

These analyses helped the team identify insights that had been missed in the focused qualitative work, and triggered the team to revisit elements of the interviews to deepen the analysis. Our data science team also helped the qualitative team understand the capabilities and limitations of emerging LLM technology and therefore develop hypotheses on which aspects of human travel expertise are unique, and which can be replicated and/or complemented by LLMs.

Following these analyses we hand-coded a "frontier dataset" to distinguish between instances of positive and negative expert practice, for use by engineers in model training and to improve the explainability of resulting product experiences.

Fig. 1. Hand-coded "frontier dataset" which labels examples of expert best practice for model training and improving explainability.

New Foundations for Human Expertise

Our framework identifies what travelers value the most about human experts when discovering and planning what to do for their next trip. Second, it evaluates which aspects of human expertise an LLM-powered "Conversational AI" can meaningfully learn from, and what remains unique to humans. Here we define a "Conversational AI" as an interface designed to simulate human-like conversation

through Natural Language Processing (NLP) and Artificial Intelligence (AI) techniques.

The framework is broken into three themes that summarize important aspects of human expertise.

- **Curating knowledge:** how human experts obtain, filter and communicate their personal knowledge about a destination with travelers
- **Personalizing interactions:** how human experts shape conversations with travelers to ensure they are surfacing and addressing their specific needs
- **Offering a perspective:** how human experts provide clear direction to travelers by sharing their unique opinion

Within each theme we identified "strategies" and related "tactics" demonstrated by human experts in response to specific traveler needs.

Curating Knowledge

Travelers turn to human experts because they seek unique knowledge that they may not be able to find elsewhere. But travelers don't want experts to share everything. They hope experts will curate their personal knowledge in a way that fits in with their specific circumstances and needs. We identified four specific needs that experts addressed.

Table 1. Curating Knowledge: Human Expert Strategies and Tactics

Traveler needs	Human expert strategies	Human expert tactics
"I want to get a sense of what something is really like"	Humans can provide details and insider knowledge by drawing on their personal experience	Identifying as local
		Detailing specifics
		Personal photos and artifacts
		Pros and cons
"I want to discover unique recommendations"	Humans can suggest off-the-beaten-path options that are not easy to find elsewhere	Hidden gems, insider tips
		Cultural references
"I want to know what all the potential options are"	Humans present a wide variety of options and	Curated lists
		Offering alternatives

	arguments within a curated range	Suggesting what's popular
"I want to learn more about the destination"	Human experts adapt relevant facts and information to add depth and context	Sharing contextual knowledge
		Sharing latest knowledge

We learned that people traveling to LA for the first time want to get a sense of what being in LA is "really like" to help them make the right decisions. Take the example of the advice shared by Jeremy, an LA local who has lived in the city for 20 years, with Donny, a 30-something dad traveling to LA with his kids.

> It is a long standing battle amongst everyone in LA who's got the best tacos. So I would humbly submit that you try to find a truck for Leo's tacos. They're my personal favorite. They're always consistent. They make these amazing Al pastor tacos. It's a seasoned pork taco. They put a little pineapple on it. It's absolutely delicious. And of course, what kid wouldn't love getting food from a truck? Hopefully that's fun for them as well.

Following the advice session with Jeremy, we asked Donny to expand on why he particularly valued the recommendation of Leo's tacos.

> I like somebody telling me, "this is my favorite". And he told me why that one was his favorite: because they add a little pineapple to the tacos when they make it. So just being able to tell me this is what I enjoy and here's why I enjoy it, I feel like that was helpful and adding a nice personalized touch by saying, this one is my favorite: this is the one that I really enjoy.

Jeremy deploys the strategy of sharing insider knowledge by drawing on his personal experience. He starts by saying "[the best taco] is a long standing battle amongst everyone in LA." Here he is subtly establishing his credibility as a local, with 20 years of experience of living in the city (and presumably eating tacos!). This imbues his advice with a specificity that is both unique to his extensive experience and embedded within local social discourse.

Second, Donny appreciates the detailed explanation of why Jeremy prefers Leo's Tacos. When he talks specifics, such as the sprinkling of pineapple, Jeremy helps him home in and appreciate aspects of the experience. This enables Donny to visualize what to expect, cutting through the paralyzing "multidimensionality" outlined by Karpik, while at the same time situating this as part of Jeremy's personal experience and perspective. This, after all, is his "personal favorite".

The value of specificity was further reinforced through the NLP analysis of the interview transcript conducted by our data science team. Similar to the qualitative analysis, the computational analysis investigates which characteristics of Jeremy's advice led to an immediate positive response from Donny. To quantify the specificity of an answer, automated tools detect the presence of named entities (geo locations, named people, restaurant names etc.) in a text and by counting such instances, we can define and evaluate at scale how detailed an expert's recommendation is. We note that the classification of the sentiment of a text is often based on the presence of sentiment-specific lexicon, in this case the prediction relies on "enjoy". Doing this at scale (enabled by the automatic tools we used) across all interviews, we discovered that during the 30 minute advice sessions, moments where the expert became specific correlated with traveler "joy".
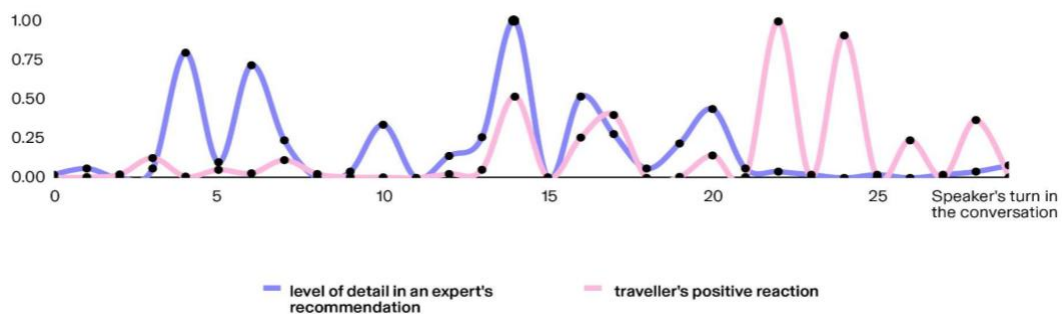
Fig. 2. NLP analysis shows experts' use of specific language partially correlates with the increase of travelers' joy score (joy is a subset of the positive sentiment).

A third and final lesson to draw from Jeremy is his referencing the relevancy of the recommendation for Donny's kids "what kid wouldn't love getting food from a truck?". Here Jeremy is calling back to the trip context Donny has shared, making the advice feel truly tailored and unique to his trip (which is the definition of personalization established by Karpik—see Personalizing Interactions below for more).

Other successful expert tactics we observed within "Curating Knowledge" included sharing personal photos and videos to showcase specific recommendations. This helped travelers visualize the experience more vividly. And seeing the expert they are talking to experiencing LA for themselves provides, perhaps, a form of vicarious enjoyment which feels more real and compelling than the photos of unknown strangers.

Elsewhere experts shared tips and tricks on how to get the most out of popular tourist destinations: where to park or which spot to have a picnic in. Travelers valued this because it was information they were unlikely to easily find, and if they did, were uncertain to trust it.

Curating Knowledge: Implications for AI Systems

The unique knowledge that human travel experts provide is anchored in their specificity as individuals. Expert knowledge is accrued through personal lived experience, creating a natural boundary between what is known and unknown. In the context of overwhelming information and the lack of an objective "right answer", these human limitations are often experienced positively by the traveler. Because it is impossible to seriously consider every option within a destination, the knowledge base of an individual expert provides a breadth of choice within a curated range, providing the traveler with a meaningful level of agency within boundaries. The near unlimited range of information available to AI systems, and the nature of their training, makes this aspect of human knowledge, especially that informed by lived experience, hard to meaningfully replicate.

Second, the specificity of lived experience also endows experts with knowledge that is unlikely to have been digitized, and therefore is not part of the data that language models are trained on. It exists within the "complex domain" of human culture, illegible to machine systems (Hoy, Bilal & Liou 2023). This "complex" knowledge can be collected from experts and translated into training datasets—as it was in this project—but this requires an intentional effort from product teams, and will not scale to the aspects of lived experience described above.

The fundamental challenge for Conversational AIs, despite their increasing capability and multimodality, is a wider epistemological weakness inherent in LLMs: the "symbol grounding problem" (Harnard 1990). As Bender and Koller explain, "language is used for communication about the speakers' actual (physical, social, and mental) world, and so the reasoning behind producing meaningful responses must connect the meanings of perceived inputs to information about that world" (2020). Because these statistical models have become deracinated from the world that produced the data to train them, their outputs necessarily lose the innate grounding of personal suggestions. And, more importantly, their training is based on the documented outputs of human culture that have been made legible via the internet and other datasets. The largely internal, mental process that an expert goes through to develop recommendations has not been digitized and therefore is not something

the models have access to. Put simply, once personal knowledge is removed from its human source, something meaningful is lost in the process.

These important limitations aside, there are aspects of the knowledge the AI systems have access to that are complementary to human expertise:

1. Data aggregation: AI can rapidly process and analyze large volumes of data from diverse sources, including travel reviews, booking patterns, and real-time information updates, at a scale far beyond human capability.

2. Global coverage: AI systems can understand and translate multiple languages simultaneously, allowing access to both local and international information about the destination.

3. Real-time information: AI can continuously monitor and integrate real-time data on weather, events, pricing, and availability across countless destinations, providing up-to-the-minute information.

4. Pattern recognition: AI excels at identifying subtle patterns and trends across massive datasets, potentially uncovering insights about travel preferences or destination popularity that might be invisible to human analysis.

5. Perfect recall: AI systems can instantly recall and cross-reference vast amounts of stored information, providing consistent and comprehensive knowledge without relying on impartial memories.

The following table demonstrates the way Conversational AIs can complement the unique human expertise of curating knowledge.

Table 2. Curating Knowledge: Human Expert and AI Complementary Strategies

Traveler needs	Human expert provides "personal insights"	Conversational AI provides "personalized information"
"I want to get a sense of what something is really like"	**Firsthand experience:** Humans can provide details and insider knowledge by drawing on their personal experience	**Aggregate for the specific:** AI can aggregate and analyze thousands of user reviews and experiences to provide a comprehensive overview of general impressions against a very specific query
"I want to discover unique recommendations"	**Uncharted gems:** Humans can suggest off-the-beaten-path options that are invisible online	**Trusted brands**: AI can identify lesser-known attractions or experiences validated by trusted sources
"I want to know what all the potential options are"	**Personal curation**: Humans present a wide variety of options and arguments within a curated range	**Comprehensive exploration:** AI can quickly compile an extensive list of all possible options from multiple sources and databases
"I want to learn more about the destination"	**Adapted to context:** Human experts adapt relevant facts and information to add depth and context	**Up to date news**: AI can provide comprehensive, up-to-date information on history, culture, climate, and current events, drawing from a vast range of continuously updated sources

Personalizing Interactions

One important expectation travelers demonstrated was that they would be able to share their travel needs discursively with the expert and receive highly personalized advice in response. We identified three strategies used by experts that made the interaction feel personalized, in response to three traveler needs.

Table 3. Personalizing Interactions: Human Expert Strategies and Tactics

Traveler needs	Human expert strategies	Human expert tactics
"I want to state my preferences in my own language"	Humans provide travelers space to state their preferences and explore them	Space for preferences
		Open ended questions and answers
		Referencing travelers' requirements
"I want a shortcut to highly relevant recommendations"	Humans ask tailored questions to address the depth of travelers needs and allows travelers to ask follow-up questions	Follow-up questions
		Assessing trip parameters
"I want to optimize the trip for me"	Humans aim to reduce friction in the itinerary	Maximizing the itinerary

A fruitful question to ask travelers following the advice session was "when did the expert not meet your expectations?". In highlighting negative experiences we learned what travelers hoped from the experience of talking to a travel expert.

For example, Yu grew frustrated with expert Ruby during their advice session because she didn't feel like she was afforded the space to truly express her preferences.

> Ruby (traveler): The conversation didn't flow as smoothly as I thought it would… I guess her answers are not open ended…you had to start another one… then start another one. I wanted it to flow from one topic to another.
> Interviewer: What could have made the conversation more flowing?
> Ruby (traveler): I think she should have asked me more questions after she's done answering so I can respond back. It should be a two way thing. But it was only me asking the questions.

Here Ruby is highlighting one of the key benefits of exploring a trip with another person: it is possible to have a back-and-forth conversation through which your requirements are explored, clarified and addressed. One of the key challenges of travel as an archetypal "singularity" (Karpik 2010) is its uncertainty. Travelers may have a general idea in their minds about what type of trip they want, but they don't fully know what's possible, nor whether those possibilities will meet their expectations in practice. As these possibilities emerge through an expert

conversation, they can visualize the trip more concretely and ask further questions, triangulating their preferences based on this feedback loop. As traveler Sheavon put it following their expert consultation, "Having a person there you get to narrow down your desires quicker. You got to do follow up questions that are answered instead of just generic websites leading you to ideas."

The most successful experts spent time at the beginning of the session asking multiple open ended questions to grasp the wider context of the traveler: who they were traveling with, past trips they enjoyed, typical activities they do on vacation, what lodgings they typically stay in. They then used this context later in the conversation to guide the traveler towards recommendations that fit with the picture they'd established, and, just as importantly, steer them away from the activities that didn't fit.

Experts explicitly referenced this context later in the session—like when Jeremy suggested Donny's kids would appreciate the Taco truck—to remind travelers they were taking their personal circumstances into account, and reinforce confidence that the item under discussion was ideal for them. They also used this context to suggest adjacent, complementary activities that might help to maximize the time spent in a particular place.

As Ruby intimates, the most successful advice sessions were not expert monologues, but those in which the traveler was afforded equal space to speak, provide feedback and clarify their own questions. This created a virtuous loop as their understanding of what was needed and possible converged through conversation. To explore and validate this phenomenon further, our data science team analyzed the extent to which both participants in the expert-traveler conversations asked a similar number of questions. This analysis further demonstrated that such conversations tend to result in travelers sharing more personal preferences, experiences, and a more positive sentiment towards the exchange. In the figure below, we highlight pair 7 where both expert and traveler are equally engaged in asking questions (both ask around 20% of questions across the duration of the interview which equates to 20 questions each in a 100-turn dialogue). When their dialogue is further analyzed using models trained for sentiment analysis and intimacy classification, the exchange proves even richer: the expert provides a safe space (measured by how positive their replies are) where the traveler admits to being a fan of visiting movie studios and consequently, receiving more personalized recommendations.
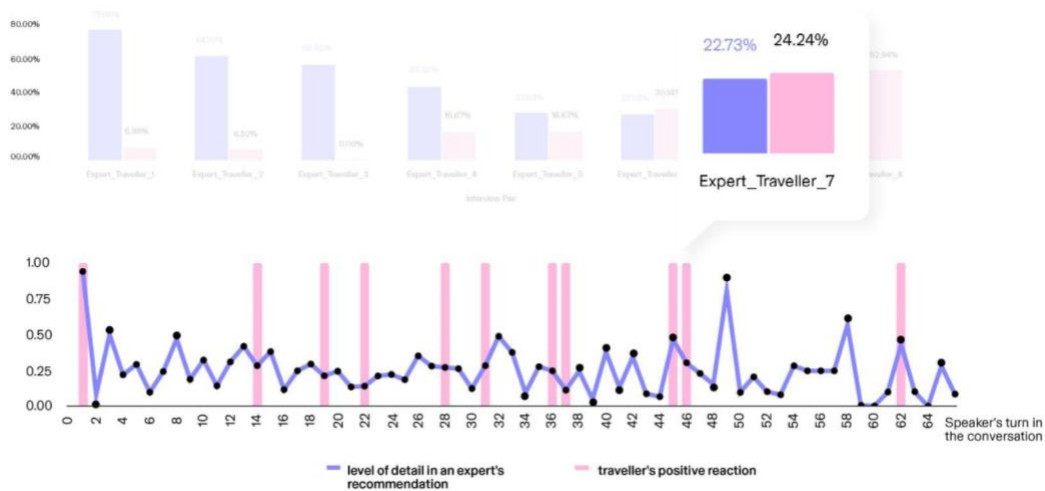
Fig 3. The first chart highlights the number of questions asked by the expert and traveler is similar (22.73%; 24.24%). The second chart shows that in the same pair there is a correlation between expert's probability of positivity (positive words used in sentences) and traveler's intimacy score (self-disclosure language used in sentences).

Personalizing Interactions: Implications for AI Systems

The personalized advice human experts provide is predicated, in part, on the social norms and cues inherent in interpersonal interaction. These norms help experts solicit information. When engaged, with intent, in a travel advice session travelers feel compelled to respond in detail to questions about their needs and context when asked directly for it by another person. Travelers also demonstrate a number of subtle cues about their engagement and interest levels through their body language, which the expert can use to "read between the lines" and interpret whether they are on the right track. A question remains around the extent to which a Conversational AI can successfully solicit and process the same fidelity of information.

Second, these social norms help skilled experts establish rapport. The most successful sessions evolved into meaningful conversations, filled with laughter, spontaneous asides and good will. When a traveler experienced rapport we observed that not only were they more likely to "open up" and share their underlying hopes and fears, but also they seemed likelier to view the advice being shared as being relevant and interesting to them. As we shall see in the final theme, "offering a perspective", because there is ultimately "no right answer" to travel, personalisation is as much about establishing confidence and trust around decisions as it is about algorithmically matching pre-existing preferences to trip items. As Forsythe explains "in everyday life, the beliefs held by individuals are modified through negotiation

with other individuals; as ideas and expectations are expressed in action, they are also modified in relation to contextual factors" (2001).

AI recommendation systems often face the challenge of creating interest and confidence in suggestions users are unfamiliar with. Data scientists call this the tension between "exploitation" and "exploration". Recommending something that is highly similar to another item that is already valued by a user is called exploitation. Exploiting known preferences provides effective personalisation in the short term, but the lack of novelty that arises can lead to repetitive experiences that erode the value of the recommendation system over time. Exploration, on the other hand, prioritizes new, unfamiliar items to a user that are at risk of not being valued at all, leading to a negative experience. But the upside of successful exploration is significant: the discovery of something invigorating and new expands personal horizons and makes the recommendation system more valuable to the user long-term.

If a Conversational AI can establish rapport like our experts, then it is more likely to convince them of these less predictable, higher value suggestions. It can explore as well as exploit. However, a big question mark remains whether a Conversational AI is able to establish this kind of rapport, both experientially (will the conversation be seamlessly intimate and responsive?) and philosophically (do travelers want to establish a relationship with a Conversational AI, and share personal contextual information with it?).

Research we've done to explore similar questions suggests that, currently, relationships with similar kinds of AI interfaces are more transactional: users want to cut to the chase when they're trying to solve a particular problem. However, there are elements that make a Conversational AI interface complementary to a human expert for personalized responses:

1. Availability and instant responses: access to travel advice anytime, anywhere, without waiting for human availability.
2. Comprehensive options: access to a huge database of trip items means that users can continue to ask for alternatives.
3. Effortless comparison and customization: easily compare multiple options and customize itineraries based on specific preferences, with the ability to quickly adjust parameters.
4. Privacy and reduced social pressure: explore travel options without the potential embarrassment or judgment that might come with asking certain questions to a human, and take as much time as needed to make decisions.

The following table demonstrates the way Conversational AIs can complement the unique human expertise of personalizing responses.

Table 4. Personalizing Interactions: Human Expert and AI Complementary Strategies

Traveler needs	Human expert provides "meaningful conversation"	Conversational AI provides "infinite iteration"
"I want to state my preferences in my own language"	**Seamless exploration:** Humans provide travelers space to state their preferences and explore them	**Unlimited prompting**: AI can enable travelers to take as much time as they need to share their thoughts, including across multiple languages
"I want a shortcut to highly relevant recommendations"	**Intuitive solicitation**: Humans ask tailored questions to address the depth of travelers needs and allows travelers to ask follow-up questions	**Endless alternatives**: AI has access to a huge database of options, enabling fast iteration to identify relevant trip items
"I want to optimize the trip for me"	**Context-specific suggestions**: Humans aim to reduce friction in the itinerary	**Unlimited filtering:** AI can integrate countless variables into suggestions to optimize itineraries

Offering a Perspective

Travelers appreciated experts who had a clear perspective, even when they didn't always agree with it. But they only listened to opinions when they trusted and related to where it was coming from. Experts exhibited three strategies that served to establish traveler trust in their opinions.

Table 5. Personalizing Interactions: Human Expert Strategies and Tactics

Needs	Strategies	Tactics
"I want to be able to contextualize recommendations so I can judge what is right for me"	Humans are upfront about their interests and biases to allow travelers to make their own decisions	Acknowledging biases
		Referencing shared interests
		Reflecting on suggestions provided
"I want to see the destination through a personal lens"	Humans let their personality and experience shine through to provide a unique lens on the destination	Personal storytelling
		Sharing opinionated suggestions
"I want to feel their excitement about the destination"	Human experts can share exciting content in an appealing format	Personal storytelling
		Positive and evocative language
		Personal destination photos

In a market in which there are "no right answers" people still seek a sense of certainty they are making the right decision. Which is why they seek the opinion of experts that they trust.

We noticed that the experts that travelers trusted the most tended to be those that they could personally relate to. When a traveler believed that an expert demonstrated a similar age, gender, interests and tastes (among other factors), they became more open to their opinions. We saw this process play out in the interaction between LA expert Jenny and traveler Christine, both of whom were female and of a similar age.

> Jenny (expert): I don't know if you've seen the photos of the Walt Disney Concert Hall. It's this amazing architectural piece. It's by Frank Gehry, who designed the Guggenheim in Bilbao. It's a really famous building, which is known for all the curvy architecture. And the concert hall in LA is really similar, so it's just like this amazing kind of metallic, silver, curved building. And even if you don't go to see a concert there, it's quite a cool building.

Subconsciously, or consciously, Jenny established credibility with Christine with these remarks. Not because of the specific recommendation in itself, but rather through what she revealed about herself in making the recommendation.

Christina (traveler): Jenny mentioned Frank Gehry, who's an architect I know. I feel like most people don't know about him off the top of their head. So that caught my ear. I was like, okay, so I feel like I trust her recommendations more because she knows who that is. I feel like if she liked that, I can picture in my head what related work of his would look like. So maybe her recommendations are in a similar vein, style wise or visually, that would align with what I'm interested in.

By mentioning Gehry specifically, Jenny helped Christina contextualize her tastes in general. This not only made that specific recommendation more valuable, but had the effect of throwing all her other suggestions—before and after—into a different, more relevant light.

Conversely when the traveler didn't find the expert relatable, this undermined even potentially relevant suggestions. Expert Leroy, intuiting that traveler Beth was seeking "off the beaten path" suggestions, proactively disclosed his perspective "If you want to go to Disneyland, I'm probably not that guy, but if you want to know about unusual stuff I am." However, following the discussion Beth wasn't convinced:

I couldn't really tell his age, but I'm assuming he's older than me. He still took into consideration that I'm younger, but then he mentioned concerts and that's pretty stereotypical. I feel like if I had someone younger, and if it was like a girl and we'd bond about shopping or something. And then she would give more of those inside scoop on where to find a vintage handbag.

Because Beth did not relate to Leroy, she read his recommendations as being "stereotypically for young people" like her, rather than emerging from his own personal experience. His opinions were therefore inauthentic.

Authenticity and self-disclosure were highly valued by travelers when conversing with experts. Our NLP analysis across conversations demonstrates that when experts shared their personal experiences and engaged in self-disclosure then travelers were more likely to use positive words in response. As seen in an earlier section, these inputs often take the form of positive feedback to the expert's advice such as "enjoying" or "looking forward" to the suggested experience made by the expert.
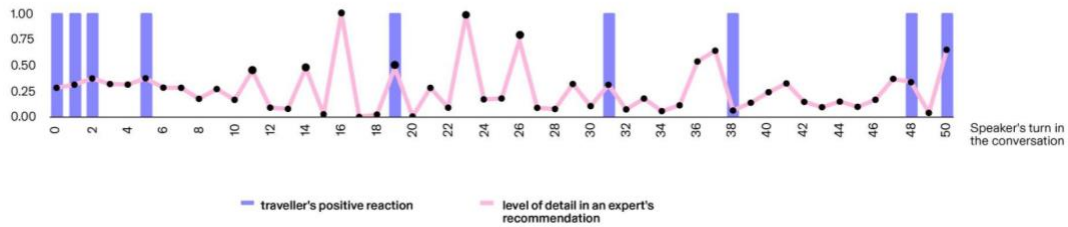
Fig. 4. The chart shows that there is a correlation between expert's intimacy score (self-disclosure language used in sentences) and traveler's probability of positivity (positive words used in sentences).

Successful experts also imbued these personal opinions with excitement, passion and positive language. Lewis, a traveler, loved Charlotte's recommendation of theme park Six Flags because of the passionate way she recounted her own experiences of the rollercoaster "it was absolutely insane!". Our NLP analysis further demonstrates how travelers respond positively when experts express excitement and passion.
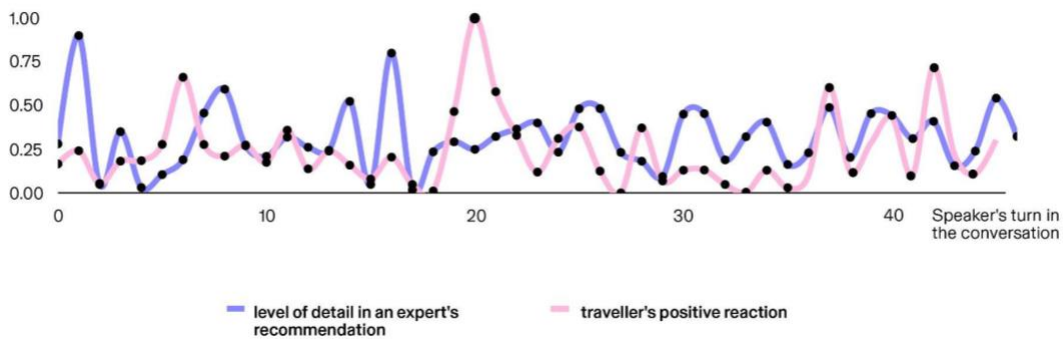
Fig 5. The chart shows that there is a correlation between expert's probability of positivity (positive words used in sentences) and traveler's probability of positivity (positive words used in sentences).

Offering a Perspective: Implications for AI Systems

The power of opinions is that they are singular; they represent the specific, authentic perspective of an individual. This situates recommendations, enabling them to be assessed in the context of their particular source. Travelers are able to pick up explicit and subtle cues about the expert to understand where the opinions are coming from, and triangulate them against their own tastes and preferences. As Forsythe summarizes: "what counts as knowledge is highly situational but is also a matter of perspective and cultural background" (2001).

It is true that Conversational AIs can represent specific characters, showcase particular characteristics, and reveal the sources of their information. But currently it is much harder for a traveler to situate themselves meaningfully against a Conversational AI because the sources of information and training are neither singular, nor embodied. Without a clear mental model of who or what they are talking to, travelers may find it hard to assess the perspective at hand.

Travelers experienced opinions as more powerful when they stemmed from the personal experiences of the expert. And if the traveler personally relates to the expert in question, then the opinion becomes even more compelling. In these moments the expert acts as a proxy, experiencing the destination vicariously on behalf of the traveler, reducing the feeling of uncertainty (Karpik 2010) in the process. Opinions that do not stem from personal experience are less authentic. Experts that did not naturally relate to the traveler in question had to defer to "second hand" suggestions, based on their general knowledge, or the experience of others. These might be highly relevant recommendations, but tended to be less trusted by the traveler.

Conversational AIs cannot experience a destination themselves, so their generated "opinions" are necessarily second hand and often a reflection of the data they were trained on. They can, of course, refer the traveler to examples of people who have experienced the destination in question (highlighting user-generated reviews, for example) but they cannot speak on behalf of those people, nor tailor the opinions of those people to the specific context of the traveler. This observation echoes the findings by NLP scholars Wu and colleagues who have highlighted that the contested reasoning capabilities of LLMs are vulnerable to degradation the greater the distance from the use case they are tested on is, as it is often the case for hyper-specific, personalized recommendations (2024).

Offering a perspective is, perhaps, the hardest type of human expertise to replicate. But there are strengths that a Conversational AI can leverage to complement and build on human perspectives:

1. Transparent source attribution: AI can clearly cite its sources for each opinion, allowing users to understand the origin and credibility of the information. This transparency can build trust.
2. Aggregation of diverse perspectives: AI can synthesize opinions from a wide range of sources, presenting a balanced view that incorporates multiple perspectives, which may be more comprehensive than a single human's opinion.
3. Personalization based on similarity: AI can identify and present opinions from travelers with similar preferences or demographics to the user, making the recommendations more relatable.

4. Quantitative backing: AI can support its opinions with relevant statistics and data analysis, adding a layer of objectivity to subjective recommendations.
5. Unbiased presentation: AI can present opinions without personal biases that might influence a human expert, potentially offering a more neutral starting point for travelers.

The following table demonstrates the way Conversational AIs can complement the unique human expertise of offering perspective.

Table 6. Offering a Perspective: Human Expert and AI Complementary Strategies

Traveler needs	Human expert provides "authentic beliefs"	Conversational AI provides "aggregated opinions"
"I want to be able to contextualize recommendations so I can judge what is right for me"	**Self-disclosure:** Humans are upfront about their interests and biases to allow travelers to make their own decisions	**Transparent sourcing:** AI can specifically reference the source on which its opinion is based upon
"I want to see the destination through a personal lens"	**Personal experience:** Humans let their personality and experience shine through to provide a unique lens on the destination	**User-generated content:** AI can surface relevant experiences from similar types of travelers
"I want to feel their excitement about the destination"	**Self-expression:** Human experts can share exciting content in an appealing format	**Aggregated experiences:** AI can demonstrate the relative excitement travelers-in-general feel towards something specific

Generalizing the Framework to Other Domains: Implications for Design and Engineering

For ethnographers exploring the value of AI within specific domains, our framework provides a useful starting point for clarifying the role Conversational AIs can play in relation to human experts. The implications of the framework span both the design of end user experience (what is increasingly termed AIUX), and the training and engineering of the underlying models themselves.

The following provides some specific principles that can be applied to any "no right answer" domain into which a Conversational AI interface is being deployed.

Knowledge: Deliver "Personalized Information" Not "Personal Insights"

Table 7. Knowledge Design Principles

Conversational AI design principle	AIUX / design implications	Training / engineering implications
Aggregate for the specific	**Do ask:** How can aggregated data address highly specific queries, without presenting it as a singular answer? **Don't ask:** How can the Conversational AI share specific, personal insights?	**Do ask:** How can personal experiences be better labeled to correspond to specific queries? **Don't ask:** How can we generate synthetic anecdotes or experiences that present a secondhand version of the real thing?
Leverage trusted brands	**Do ask:** How can we clearly attribute information to specific trusted sources while maintaining a seamless user experience? **Don't ask:** How can the Conversational AI curate recommendations itself?	**Do ask:** How can different sources be labeled to correspond to specific queries and user types? **Don't ask:** How can the Conversational AI access firsthand knowledge that hasn't been shared online?
Enable infinite exploration	**Do ask:** How can we provide a balance between curated recommendations and user-driven exploration? **Don't ask:** How can we arbitrarily create limited lists that reflect the limits of individual knowledge?	**Do ask:** How can different options be labeled for quality and relevance against specific contexts? **Don't ask:** How can we artificially expand the dataset to cover all possible scenarios without proper validation?
Surface live information	**Do ask:** How can live access to different feeds enable cutting edge, up to date information? **Don't ask:** How can we assume what live information is relevant to the individual without clear signals?	**Do ask:** How can we implement real-time fact-checking mechanisms to ensure the accuracy of live information? **Don't ask:** How can we prioritize recency over accuracy or relevance in information retrieval?

Interaction: Provide "Infinite Iteration" Not "Meaningful Conversation"

Table 8. Interaction Design Principles

Conversational AI design principle	AIUX / design implications	Training / engineering implications
Encourage additional prompting	**Do ask:** How can we design an interface that encourages free-form expression and prioritizes questions over answers? **Don't ask:** How can we replicate the situated understanding of cultural context that human experts bring to language interpretation?	**Do ask:** How can we improve NLP to accurately interpret preferences that are expressed colloquially? **Don't ask:** How can we program the AI to have personal travel experiences in different cultures to make it more relatable?
Leverage the signals	**Do ask:** How can we present a wide range of options while quickly narrowing down to the most relevant based on user interaction? **Don't ask:** How can we mimic the intuitive ability of human experts to read between the lines and infer unstated preferences?	**Do ask:** How can we optimize algorithms to rapidly filter and prioritize options based on user inputs and behavior patterns? **Don't ask:** How can we replicate the human ability to draw on experiences for nuanced recommendations?
Offer unlimited filtering	**Do ask:** How can we create an interactive interface that allows users to easily adjust and visualize different optimization parameters? **Don't ask:** How can we simulate the empathetic understanding that human experts use to balance conflicting preferences in trip planning?	**Do ask:** How can we develop algorithms that consider multiple variables simultaneously to create truly optimized itineraries? **Don't ask:** How can we program the AI to have the contextual awareness and flexible problem-solving skills of seasoned travel experts?

Perspective: Share "Aggregated Opinions" Not "Authentic Beliefs"

Table 9. Perspective Design Principles

Conversational AI design principle	AIUX / design implications	Training / engineering implications
Provide transparent sourcing	**Do ask:** How can we design an interface that clearly presents the sources and context for each recommendation? **Don't ask:** How can we replicate the human expert's ability to intuitively share personal biases and interests?	**Do ask:** How can we implement a system to accurately track and cite sources for all information used in recommendations? **Don't ask:** How can we program AI to develop its own interests and biases based on a particular perspective?
Match to user generated content	**Do ask:** How can we create a user interface that effectively presents and filters user-generated content based on traveler similarities? **Don't ask:** How can we mimic the unique personality and experiential narrative that human experts bring to their recommendations?	**Do ask:** How can we develop algorithms to match user profiles with the most relevant user-generated content? **Don't ask:** How can we give the AI its own personality and set of travel experiences to draw from?
Aggregate sentiments as well as information	**Do ask:** How can we design a way to present aggregated excitement levels visually or interactively to engage users? **Don't ask:** How can we make the AI genuinely feel and convey personal excitement about a destination?	**Do ask:** How can we develop sentiment analysis tools to accurately gauge and represent traveler excitement levels from various data sources? **Don't ask:** How can we program the AI to have authentic emotional responses to destinations?

Conclusion: Exploring the Space between "Autonomous Agent" and "Extended Self"

The debate about the future of AI computing interaction is divided between two visions. The first, articulated by OpenAI CEO Sam Altman, is that of AI as a "super-competent colleague that knows absolutely everything about my whole life, every email, every conversation I've ever had, but doesn't feel like an extension" (Varanasi

2024). This is the classic vision of AI as an agent with an independent identity and the capacity to act autonomously.

The second vision is articulated by Apple CEO Tim Cook. For Cook, the "Apple Intelligence" platform is about "empower[ing] you to be able to do things you couldn't do otherwise. We want to give you a tool so you can do incredible things." (Indian Express 2024). This is the more incremental vision of AI as an embedded feature that gives users new capabilities, in line with Steve Job's vision for computing as providing a "bicycle for the mind". In this view AI is experienced as an extension of the self, rather than as an autonomous, separate entity.

These distinct visions provide us with a useful spectrum for thinking about the future of expertise. On one end we have the OpenAI vision of the Agent and, on the other, the Apple vision of the Tool. Applying these two concepts to a conversational experience in travel, the Agent approach benefits from its autonomy, and is therefore perhaps better positioned to challenge, inspire and act on behalf of travelers as a human expert would. On the downside, the inevitable personification of such an Agent, especially when experienced through a conversational interface, may trigger expectations of human expertise, and the "personal insights", "meaningful conversations" and "authentic beliefs" that we have discovered people value. The fact that these expectations are unlikely to be satisfied for the reasons outlined above may create disappointment and confusion.

The more conservative vision for AI articulated by Apple will not trigger these unrealistic expectations, but could suffer from the opposite problem. If an AI is only experienced as an incremental Tool in the context of an existing product journey, then user expectations will be constrained by the pre-conceived limits of that specific context.

Technologist Matt Webb nuances the emerging AI landscape further by adding "Chat" as a third pole, defined as the capacity to "talk to the AI as a peer in real-time" (Webb 2024). For Webb, Tools, Agents and Chat are distinctive user modalities for interacting with AI, each enabled by a different aspect of the underlying technology. An AI that works alongside a user as a "Co-pilot" incorporates elements of all three modes.

Fig. 6. Matt Webb distinguishes between different AI user experiences, enabled by three core capabilities.

What this increasingly complex picture surfaces is the lack of adequate mental models to help people successfully navigate these emerging modalities, and, by extension, what to expect from AI-driven expertise. This presents a significant opportunity for ethnographers. The travel case study has helped us identify the underlying, complementary qualities that AI can provide alongside human expertise. What is less clear, however, is how these complementary qualities should be experienced in domains outside of travel, including through what modality.

The design principles outlined above are useful starting points for product teams seeking to build new conversational experiences across consumer domains, especially those for which there is "no right answer." However, our experience is that the implications for user experience (AIUX) and engineering will be specific to the context in question. This requires teams of ethnographers and data scientists (using combined methods like those outlined in this paper) to carefully explore the cultural and behavioral nuances of the domain they are designing for, and convert this understanding into outputs that are legible to engineers. As with any type of product, the difference between success and failure will be defined by the value it provides to the end user, not the raw potential of the underlying technology.

About the Authors

Tom Hoy is Partner and co-founder of Stripe Partners. His expertise lies in integrating social science, data science and design to unlock concrete business and product challenges. The frameworks developed by Tom's teams guide the activity of clients including Apple, Spotify and Google. Prior to co-founding Stripe Partners, Tom was a leader in the social innovation field, growing a hackathon network in South London to several hundred members to address local causes.

Janneke Van Hofwegen is a UX Researcher on the Real World Journeys team on Google Search. She employs ethnographic and quantitative methods to learn how people are inspired by and decide to embark on experiences in the world. She holds a PhD in Linguistics from Stanford University. Her linguistics work focuses on understanding human behavior and stylistic expression through the lens of language and group/individual identity.

References Cited

Bender, E., & Koller, A. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185-5198). Online: Association for Computational Linguistics.

Buchanan, B. G., R. Davis, and E. A. Feigenbaum. 2006. "Expert Systems: A Perspective from Computer Science." In *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. Anders Ericsson, Neil Charness, Paul J. Feltovich, and Robert R. Hoffman, 87-103. Cambridge: Cambridge University Press.

Chang, L. 2023. "Most Visited American Cities." Condé Nast Traveler. Available at: https://www.cntraveler.com/story/most-visited-american-cities.

Confente, I. 2015. "Twenty-Five Years of Word-of-Mouth Studies: A Critical Review of Tourism Research." *International Journal of Contemporary Hospitality Management* 17(6): 613-624. https://doi.org/10.1002/jtr.2029.

Dreyfus, H. L. 1987. "From Socrates to Expert Systems: The Limits of Calculative Rationality." *Bulletin of the American Academy of Arts and Sciences* 40 (4): 15-31.

Forsythe, D. E. 2001. *Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence.* Stanford: Stanford University Press.

Frey, C. B., and M. A. Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254-280.

Harnad, S., 1990. The symbol grounding problem. Physica D: Nonlinear Phenomena, 42(1-3), pp.335-346.

Hawkins, A. J. 2024. "Google AI Gemini Travel Assistant Now Handling Hotel Bookings." The Verge. Available at: https://www.theverge.com/2024/5/14/24156508/google-ai-gemini-travel-assistant-hotel-bookings-io.

Herbold, S., A. Hautli-Janisz, U. Heuer, Z. Kikteva, and A. Trautsch. 2023. "A Large-Scale Comparison of Human-Written Versus ChatGPT-Generated Essays." *Scientific Reports* 13 (1): 18617.

Hoy, T, Bilal, I.M, and Liou, Z. 2023. "Grounded Models: The Future of Sensemaking in a World of Generative AI." *EPIC Proceedings* 2023: 177-200.

Karpik, L. 2010. *Valuing the Unique: The Economics of Singularities*. Princeton, NJ: Princeton University Press.

"Tim Cook Explains Why Apple Calls Its AI 'Apple Intelligence'." 2024. Indian Express. Available at: https://indianexpress.com/article/technology/artificial-intelligence/tim-cook-explains-why-apple-calls-its-ai-apple-intelligence-9389541/.

Jacobsen, J., and Munar, A.M. 2012. "Tourist Information Search and Destination Choice in a Digital Age." *Tourism Management Perspectives* 1: 39-47.

Lowery, A. 2023. "How ChatGPT Will Destabilize White-Collar Work." The Atlantic. Available at: https://www.theatlantic.com/ideas/archive/2023/01/chatgpt-ai-economy-automation-jobs/672767/.

Mountcastle, B. 2024. "Using ChatGPT for Travel Planning." Forbes. Available at: https://www.forbes.com/advisor/credit-cards/travel-rewards/chatgpt-for-travel-planning/.

Pierce, D. 2023. "The AI Takeover of Google Search Starts Now." The Verge. Available at: https://www.theverge.com/2023/5/10/23717120/google-search-ai-results-generated-experience-io.

Varanasi, L. 2024. "Sam Altman Wants to Make AI Like a 'Super-Competent Colleague That Knows Absolutely Everything' About Your Life." *Business Insider*. Available at: https://www.businessinsider.com/sam-altman-openai-chatgpt-super-competent-colleague-2024-5.

Webb, M. July 19, 2024. "The AI Landscape." *Interconnected*. Available at: https://interconnected.org/home/2024/07/19/ai-landscape.

Wu, Z., L. Qiu, A. Ross, E. Akyürek, B. Chen, B. Wang, N. Kim, J. Andreas, and Y. Kim. 2024. "Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Volume 1: Long Papers), 1819-1862. Mexico City, Mexico: Association for Computational Linguistics.