Is artificial intelligence coming for information designers' jobs?

Karen A. Schriver, PhD
May 1, 2023

Note: This pre-publication is an editorial for the *Information Design Journal* (IDJ), 28.1 (2023).
Amsterdam: John Benjamins Publishing Company.

Karen A. Schriver

Editorial

Is artificial intelligence coming for information designers' jobs?

1. Are we ready for AI chatbots?

As professionals whose business it is to integrate word, image, and type, we need to question our readiness for new uses of artificial intelligence on the horizon. I'm talking ChatGPT (now powered by the GPT-4 language model) and the impact of large language models (LLMs) on information design. A key feature of GPT-4 (released in March of 2023) is its ability to take in both images and text to generate answers in text. Earlier versions were already reasonably good at tasks like translation and summarization, but the latest version can perform tasks such as writing business documents, academic articles, generating art, music, 3-D images, animations, and even games.

GPT-4—trained by developers at OpenAI—is built on vast amounts of data scraped from the internet and elsewhere—ranging from news articles, images, artwork, and encyclopedias to social media posts and online forums. Although its comprehensiveness is impressive and better than GPT-3.5, part of its database is still loaded with bias and misinformation, raising serious concerns about the ethics that underlie its feedback.

Additionally, it still falters in human-level professional performance, leaving important things out and making up other things (called hallucinations). While it's full of information and predictive power, it lacks rhetorical sensitivity, and according to its founders, "makes reasoning errors" (OpenAI, 2023, p. 10).

As people who mediate human experience through symbols and signs, we information designers need to think seriously about how advances in AI will change our jobs, our image of what we do, and perhaps our field.

2. Jumping into the AI fray as information designers

This issue of *IDJ* is not focused on AI and chatbots, though the first article is about deep learning. Still, the topic of deep learning and GPT-4 is timely—inviting us to think about how AI might help or hinder the authors of this issue (and the contributions of future authors). As these large language models become more pervasive for writing and designing documents as well as for judging them, we need to understand the assumptions that underlie them. As you read this issue, consider how the authors' writing and design process might be different with AI. What might AI contribute to our understanding of the topics the authors discuss? Would AI change their literature reviews, data analysis, or conclusions?

Let's first get a sense of what this issue is about. I'll offer a summary of each of the four articles and then reflect on how AI might influence them.

3. Featured in this issue

3.1 A research article on using AI to create a readability formula

In 'Language independent optimization of text readability formulas with deep reinforcement learning', Moghaddam and Ghayoomi present a case in which they used AI and deep learning to assess text difficulty in English and Persian. Their goal was to update several well-known readability formulas and generate a new AI-driven formula. In doing so, their formula extracted text features from an English and Persian dataset.

Their formula looked at text features beyond those counted in the traditional Flesch readability formula (the one found in *MS Word's Editor).* While most formulas typically calculate word length and sentence length, theirs included lexical diversity, prepositions, nouns, verbs, adjectives, adverbs, determiners, and other linguistic features. They also refined their formula by using a deep reinforcement learning model. After several rounds of optimizing the formula's parameters, the authors' showed improved accuracy in distinguishing among beginning, intermediate and advanced texts in English and Persian.

The model proposed by the authors is intriguing in concept because it promises to use deep learning methods to perfect readability formulas. The AI model extracts linguistic characteristics of texts beyond the number of syllables or the number of words per sentence. It looks for text features that presumably contribute to readability. It then uses the features to train the AI model on what to look for in target texts. After training, the model was able to sort the target texts into beginning, intermediate, or advanced—making use of the linguistic features. The authors' model was successful in classifying the texts they studied into distinct groups in two languages.

Comment

Skeptical information designers will expect much more research on texts for children and adults before models such as this one could be used productively. Crucially important is establishing a reliable measure to judge what is meant by comprehension.

Taking a dataset to guide training and using a modified version of an old formula may generate a better formula, but do the results benchmark readers' comprehension? Would the formula work in the wild? What precisely are the linguistic characteristics of a beginner text or an advanced text? Ideally, we would begin the design process by having people who read at different reading levels show us what is basic or advanced. We want human performance to reveal textual, graphic, or typographic quality. And if the text is for children, children's behavior not teachers' opinions should decide what is understandable. In this way, we can use AI to calibrate a valid standard for reading levels based on human comprehension and use.

3.2 A research article on choosing bilingual typefaces

In 'Beyond intuition: An empirical study of typeface selection in a bilingual context', Li and Westland present a method for choosing type in contexts that present the same content in two languages. Their goal is to help designers convey a coherent visual message in contexts such as reading a restaurant sign in both Chinese and English.

They assert that well-chosen bilingual typefaces can express meaning by evoking emotions or revealing personalities. Rather than classifying the coherence of typefaces based on unique characteristics of letterforms, they want to capture the hybrid emotive meanings of a typeface—such as "organized while slightly contemporary." They propose an idea for finding the visual similarity among typefaces based on the target audience's judgements. Participants in their study rated bilingual typefaces for consistency or contrast.

If the authors are right, even non-designers can use their method to select typefaces. They propose that people creating AI-based designs with big data, for example, could use the method to generate typographic suggestions for text-to-design applications. They contend that all one needs to know is the content or the idea one is writing or designing about. Presumably, the method will be simple to use once programmed.

Comment

The authors argue that too many designers make typographic decisions based on instinct and personal preference. With this view, designers treat the audience as a passive spectator of their informed choices.

Interestingly, there is a parallel between designers who work by intuition and AI researchers who do the same, drawing on training sets in inadvertently unthinking ways. Designers and AI researchers need to think about whose assumptions, needs, values, language preferences, and cultures are both represented and promoted in their work. The strength of the authors' approach is its grounding in human experience; a weakness is the limited number of typographic pairs they tested. Of course, AI could help in testing many more pairs.

3.3 A research review on the role of typography in procedural information

In 'Type does matter! A systematic quantitative literature review', Schmid, Carim, Sargent, and Falla evaluated the recent advice about typography for electronic documentation in the high-risk fields of aviation and medicine. They studied typographic guidelines, mainly for procedural documents—manuals, maps, charts, checklists, performance aids, or documentation samples—to figure out how the advice improves or impairs legibility.

Schmid and his colleagues underscore the need for improved legibility of electronic documentation—especially in professional contexts that depend on safety, training, and accountability. They analysed 65 publications to determine the level of detail they offered in recommendations about the legibility of type. They rated the publications according to their depth of coverage (none, shallow, and in-depth) and for the presence of ideas about typography for digital or paper documents.

They found that most publications in medical and aviation domains had a rather shallow treatment of typography. Typographic conventions were often based on inherited misconceptions. Moreover, publications gave the same advice for paper and online without considering how electronic presentation may need to be different.

Strangely, most of the advice about type was created by domain experts in aviation or medicine. Of the 65 articles the team reviewed, only one had been co-authored by someone with experience in typography. Not surprisingly, subject-matter experts tended to recommend the standard typefaces bundled with the Windows™ operating system.

Some articles drew comparisons between two typefaces displayed at the same size, but that differed in character height. This led to unconvincing results. And the most egregious omission in the typographic guidelines was that the people and their context for using the documentation were ignored.

Comment

One of the things this paper tells us is that the literature on the legibility of typography for documents about aviation or medicine has not progressed much since the early 1990s. For example, the authors found a lot of vague advice, such as "avoid typefaces containing characters that are too similar to one another". It would have been more informative to tell users which characters are most at risk of being misread. In particular, the lower-case L, upper-case I, and the number 1, but also the characters C, c, e, O, o. They also found that none of the publications addressed the potential problems of emphasizing words in bold, which may reduce the negative space within letters, thus making them more difficult to read. Now that GPT-4 can analyze images, it could be taught to flag typographic elements that contribute to illegibility.

3.4 A personal reflection from an information designer

A final contribution to this issue comes from Pettersson, who offers a personal reflection on his career. He catalogs his journey as a Swedish information designer over many years and describes how the field has changed since he started working.

In some ways, Pettersson's story reads like the history of technology. He recounts his early days of working with primitive versions of audio-video equipment and the trials and tribulations of working for big corporations and academic institutions.

He says that many publishers he worked for in the 1960s–1970s feared the coming of the *electronic revolution*. They believed that new media would soon replace books, that paper was dead. So they slashed their editorial staffs. Not only did designers come to dislike the idea of new media, but they grew suspicious of employer's motives.

Interestingly, many of today's writers and designers are wondering if AI is coming for our jobs. Once again, we find ourselves chasing new technologies to master. In the case of AI, however, there may be no mastering. But certainly we can learn to use it productively so we can focus on more sophisticated tasks.

Comment

In his article, Pettersson narrates his experiences with books, especially the design of textbooks for Swedish students and guidebooks for their parents. He tells us that today his grandchildren use no textbooks since everything is online. I suppose his next job will be to ensure that his grandkids are not plagiarizing other peoples' work by using ChatGPT-4 to draft their papers for them.

4. Closing remarks

Research has shown that readability formulas (the topic of paper 1) are rather weak tools for assessing text difficulty (Schriver, 2017). Reading researchers developed traditional formulas based on English texts and used a rather limited set of parameters for defining a grade level for a text (e.g., word and sentence length). There is reason to believe that within a few years AI tools will derive better text analysis software because they can analyze a range of linguistic text features (e.g., words, grammar, syntax, structure) at scale. And importantly, they can evaluate texts across a range of languages looking for critical patterns that predict clarity and cohesiveness.

As information designers, we need to be watchful of the use of AI and deep learning models claiming to evaluate text quality. Because we stand between visual-verbal content and the reader or user, we must ensure that AI does not lead organizations or the public to a false sense of what text quality means. A text free of grammatical or syntactic errors does not mean it is a good text for a reader. We also need to inspect the content that AI generates and scrutinize its output for truthfulness, omissions, or slanting of content. Whose story gets told and whose gets left out is as important as a story being well told.

It is not clear that those who train the AI models have considered the ethical dilemmas their natural language generation capabilities may pose. It is not clear they will generate texts that are responsive to peoples' needs. It seems likely that within the next year advances in neural networks and AI deep learning techniques will produce models that both generate and evaluate visual and verbal texts. We need to be vigilant in questioning what metrics are brought to bear to measure text quality and whose texts are being used for training the models.

Information designers create a lot of texts aimed at both understanding and use. Certainly AI may have a disruptive impact on this work. For example, it appears that procedural instructions will be the low-hanging fruit for these large language models to automatically generate.

It's not just procedural documents though. An MIT economics team recently found that ChatGPT can help professionals write business documents faster and of higher quality (Noy & Zhang, 2023). They studied 444 college-educated professionals who wrote press releases, short reports, emails, and analysis plans in one of two conditions: with the help of ChatGPT or without its help.

They found that professionals in the ChatGPT condition spent significantly less time writing, increasing their productivity. Professionals with help from AI spent less time brainstorming and generating a rough draft (offloading the work to ChatGPT), but more time editing. Still, their overall writing time from start-to-finish was less. A group of judges (who were blind to which texts were written with or without ChatGPT) rated the AI-assisted texts higher than those written without help. The researchers—who focused on automation and productivity—said this about ChatGPT and the employment of writers:

> A potent generative writing tool like ChatGPT might entirely replace certain kinds of writers, such as grant writers or marketers, by letting companies directly automate the creation of grant applications and press releases with minimal human oversight. *This might not increase the quality of the resulting written output but would let companies save on wage costs by eliminating human labor* [italics added]. Alternatively, a tool like ChatGPT could substantially raise the productivity of grant writers and marketers, for example, by automating relatively routine, time-consuming subcomponents of their writing tasks, such as translating ideas into an initial rough draft. In this case, demand for these services could expand, resulting in higher employment and wages as well as greater productivity for companies and cheaper products for consumers. (Noy & Zhang, 2023, p. 1)

The authors also contend that professionals who are poor writers are helped more by ChatGPT. Their writing improved more than did stronger writers. For high ability writers, their process got a lot faster. Still, two weeks later there was no improvement in job satisfaction for either group.

AI models may seem like innocent algorithms chugging away at refashioning the data they acquire during training. Aye, but there's the rub: It's the training data we need to worry about. Some have called these language models *stochastic parrots* (Bender et al., 2021) because they parrot back what they've learned—creating accurate and inaccurate output. The good, the bad, and the ugly.

According to MIT's *Technology Review Insights* (2023), there are two things missing from AI algorithms today. The first is accountability for the decisions made by these systems. That is, often we do not know why a program does what it does and how it draws its conclusions. They acknowledge that historical social inequities are baked into the raw data. The second thing we don't know is whether organizations should trust the decisions bots make and the advice they give to their customers and employees.

Clearly AI can be used for nefarious purposes. As Frascara (2004) reminds us, "design is never neutral" (p. 119). Let us keep our eye on issues of ethics and human values as we move forward.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, Mar. 3-10). On the dangers of stochastic parrots: Can language models be too big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. https://dl.acm.org/doi/10.1145/3442188.3445922

Frascara, J. (2004). *Communication design: Principles, methods, and practice*. New York: Allworth Press.

MIT Technology Review Insights. (2023, Feb. 14). Deploying a multidisciplinary strategy with embedded responsible AI. *MIT Technology Review*. https://www.technologyreview.com/2023/02/14/1066582/deploying-a-multidisciplinary-strategy-with-embedded-responsible-ai/

Noy, S., & Zhang, W. (2023, Mar. 10). Experimental evidence on the productivity effects of generative artificial intelligence. https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1_0.pdf

OpenAI. (2023, Mar. 27). GPT-4 *Technical Report*. https://arxiv.org/pdf/2303.08774.pdf

Schriver, K. A. (2017, Dec.). Plain language in the United States gains momentum: 1940–2015. *IEEE Transactions on Professional Communication,* 60 (4), 343–383. http://ieeexplore.ieee.org/document/8115322/ Available: https://www.plainlanguage.gov/media/Schriver%20Plain%20Language%20in%20US%20Gains%20Momentum%201940_2015%20Draft.pdf