Chatbot Quick Facts

Chatbot Workflow Steps

1. A book (in XHTML/EPUB format) is selected for inclusion in the chatbot.
2. The book is parsed using a user interface.
3. Vector embeddings are generated for the book.
4. Vector embeddings and corresponding book contents are stored in a database.
5. The website is updated to reflect the book's chatbot-readiness.

Author Information

When we enable the chatbot for a particular book, the author may ask some variation of, "Is my book being used to train LLMs?" Importantly, we are *not* doing this–**the only generative AI used is the base ChatGPT model**. According to OpenAI's Enterprise Privacy Policy, they don't train their models on any Hopkins Press data, and all data ownership is retained by Hopkins Press.

The books *are* being embedded, which is an AI process that converts chunks of text into mathematical representations. A computer is able to compare these mathematical representations: the more similar they are, the more similar the corresponding words are. **Book content is only being used in an AI (Natural Language) search** to pull the most relevant sections of a book, which is then cleaned up by ChatGPT and presented to the user.

The user will never have access to the book content itself. Their questions are answered as to whether the book has the requested information. For instance, asking "How can AI be used to improve educational outcomes in the classroom?" does not give actual strategies in the book, but instead directs the user to the most relevant chapters.

In addition to the harm reduction above, the author may *want* their book included for the following reasons:

- **Participation in innovative project**: Recent developments in AI have been transforming the information landscape. Many of these changes, like AI search, are inevitable, and this would be an opportunity for authors to experience it from a trusted source like Hopkins Press.
- **Improved discoverability**: Academics won't have to rely on keyword searches to find the best content. Instead, they can ask their research question directly to the chatbot and immediately find the most relevant chapters of the book.
- **Greater customer trust**: By providing a potential customer with insights into specific sections of the book, customers who were on the fence may decide it *is* the right book. Further, customers who may have bought the book only to find it did not meet their needs can discover this before buying the book, preserving trust for future sales and interaction with Hopkins Press.
- **Marketing insights**: Customer queries can provide Books Marketing with greater insight on how customers are searching for and determining whether to purchase books. (Note: this isn't implemented yet, as it raises privacy questions.)

Further, having the book embeddings opens up the book to more products:

- **Inclusion in book-azines**: Embeddings can be used to identify and associate similar book topics/sections for inclusion in a book-azine, leading to further discoverability of the book.
- **Full-text AI search**: The underlying vector embedding technology can easily be extended to a full-text search for books, allowing users to make requests like "Show me books on the impacts of the civil war" and see *all* relevant books from the Press.
- **Cross referencing on similar books**: Rather than relying solely on human selection of books, embeddings can be used to identify "Related Books" based on similarity of the texts.