AI in Libraries: Seven Principles

May 2020

KB nationale bibliotheek

Seven Principles for AI in Libraries

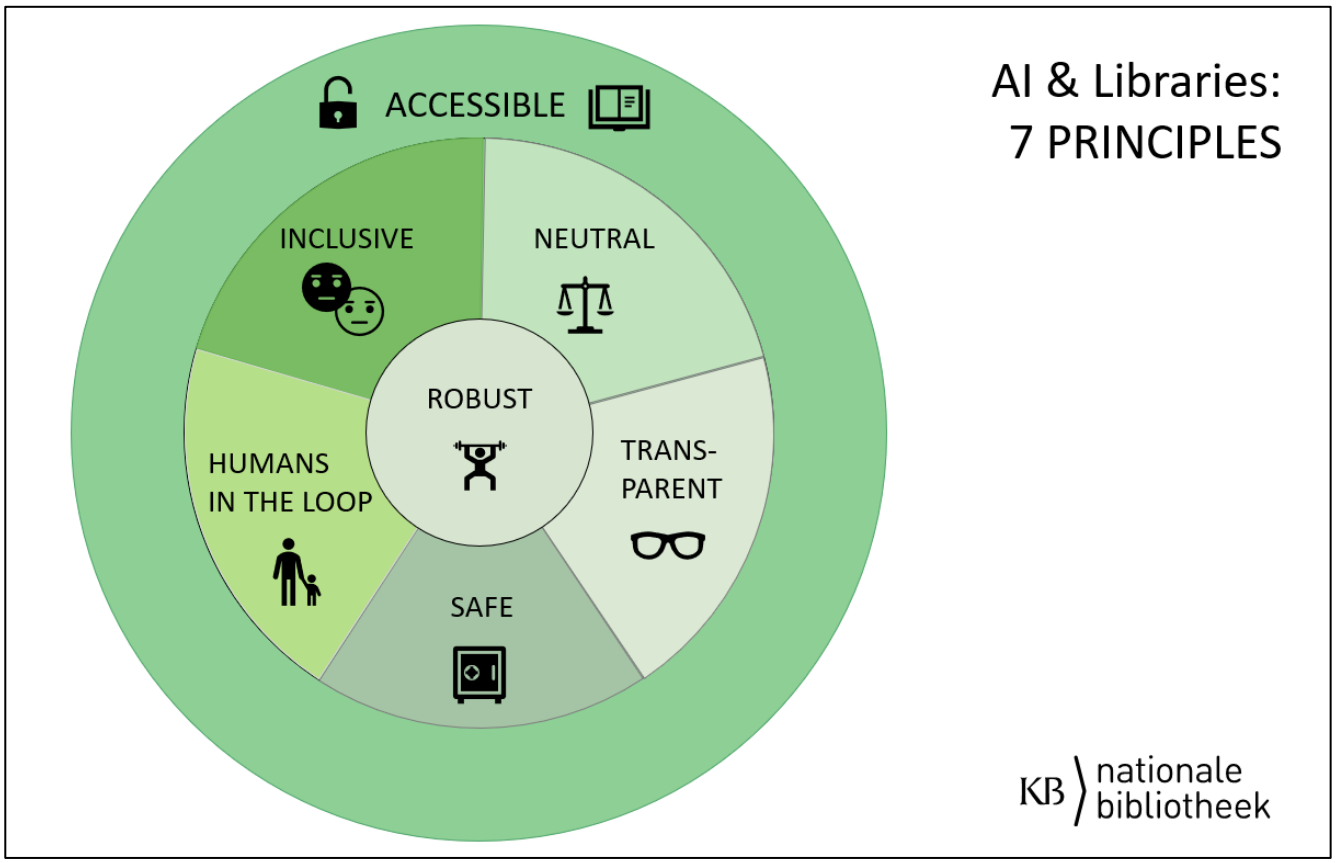
Artificial Intelligence (AI) plays an increasingly important role in reading, learning and research, the core activities of the library. The National Library of the Netherlands (KB) appreciates the great possibilities that AI offers to help us accomplish our mission to make our users smarter, more creative and more skilled.

At the same time we do not close our eyes for unforeseen and unwanted side effects of AI, which may raise ethical questions, perhaps more than other forms of digital transformation.

As National Library we feel we have a responsibility in addressing these issues. We have formulated seven principles we feel any AI-application in the context of the Library should comply with. Seven principles that takes the potential and the benefits of AI as starting point, and at the same times give guidelines to avoid ethical pitfalls.

KB, National Library of the Netherlands
Jan Willem van Wessel
janwillem.vanwessel@kb.nl

Our starting point for defining AI principles is the central premise of the UN's *AI for good* initiative, stating that AI can and must be used to make the world a better place. To define a better world, the United Nations have formulated 17 Sustainable Development Goals (SDG). Many initiatives and projects, including *AI for good*, are guided by these SDGs.



1. Accessible

The library uses AI primarily to make information accessible to the public and to promote (digital) literacy of all citizens.

First and foremost, we use AI to help libraries accomplish their mission to make as much information as possible accessible for our current and future users to read, learn and research. Our main task is to preserve information and make it accessible and discoverable. Applying AI in libraries should support the United Nations' sustainable development goals (SDG), specifically SDG 16.10 *providing public access to information* and SDG 4.6 *promoting literacy and numeracy*.

Using AI, literacy of people can be improved with individually tailored learning and reading programs.

Furthermore, AI makes it possible to ask questions in spoken, natural language.This makes information better accessible to more people. AI can also read aloud texts to people who do not (yet) read well. AI brings expensive production of audiobooks within reach by enabling automatic conversion of written texts to spoken word.

2. Robust

We only work with AI applications that are robustly designed and developed and that are reliable in use.

AI applications are software applications. Its robustness is determined by the quality of design, realization, testing, maintenance and documentation. AI suppliers must use proven methods, standards, norms, frameworks and certifications for software engineering, supplemented with new standards for machine learning algorithms and data sets for training and testing.

An AI application must meet high standards. AI cannot be functioning 99% of the time if critical decisions rely on it. AI cannot give answer A the first time and answer B the next time, if nothing that is relevant to the outcome as changed.

It may be acceptable for book recommendation software to come up with five suggestions, of which four are correct. But it becomes problematic if the same AI software is used to recommend scientific articles and that fifth recommendation contains the latest research.

3. Inclusive

We only develop and use data sets and AI applications that are inclusive.

We realize that all AI is inherently susceptible to bias. We use our expertise to make users aware of biases that AI-systems and their output may have. More importantly, libraries should play an active role in creating inclusive AI by making datasets available for training and testing AI-applications that are either unbiased, or explainable biased towards age, ethnicity, religion, gender identity, sexual orientation, origin and political preference.

As KB, we provide datasets for research into and development of AI. We make sure that these datasets are as little biased as possible with regard to age, ethnicity, religion, gender identity, sexual orientation, origin and political preference.

In practice, however, this is hard. In fact, in our (digital and physical) depots there are no data sets that are 100% unbiased, if only because of changing views on collection policy over the years.

Therefore, more important than preventing bias, is to know where and to what extent bias occurs, so that it can be eliminated or compensated, thus maintaining inclusivity.

4. Neutral

We do not develop or use AI applications that actively aim to manipulate people's behavior or thinking.

While AI systems are designed to support humans at making decisions and choices, this may never be misused by steering users into directions that they would not choose outside the context of AI.

Moreover, partiality in commercial or political choices and decisions undermines diversity. Like wth inclusivity, the neutrality of AI systems can be influenced by targeted (training) data and algorithms.

Here, the library must assume their role as a neutral expert and guide their users and provide neutral advice.

5. Humans in the loop

We only develop and use AI applications that have, by design, at crucial points, some form of monitoring by a human. 'No human no AI.'

AI serves to support human activities but should not act as an independent decision system. AI can perform many tasks independently, but must at crucial points be monitored by people who ultimately decide.

Human monitoring does not necessarily have to involve training of AI or checking all output of the system. After all, one of the goals of AI to make this unnecessary. It concerns the critical testing and assessment of AI applications, both in terms of the principles formulated here, and in areas where AI is far from mature, such as causal inference and social intelligence.

6. Transparent

Preferably and where possible, we only develop and use AI applications whose algorithms, training data and method are transparent.

AI does not have to be a 'black box' of which even the designer cannot explain how the AI arrived at a certain conclusion. Transparency can be achieved by applying insights from Explainable AI (XAI). An important premise of XAI is the "right to an explanation," which is an individual's right to have an explanation of how an AI system arrived at a certain conclusion, especially if that conclusion has financial, legal, or social implications.

To this end, XAI is developing methods and techniques that enable human experts to explain the results of an AI application, even if full transparency is not always possible due to the complexity or confidentiality of the algorithm.

7. Safe

We only develop and use AI applications of which we are sure they respect the privacy of our employees and users.

Safe AI is AI of which we are certain it respects the privacy of our employees and users. This principle does not apply exclusively to AI. There is extensive privacy legislation, especially in the context of IT systems, that applies to AI-systems as well.

There is substantial social concern about privacy in relation to AI. From facial recognition of people who think they move anonymously in a public place, to combining data from various big (user) data collections for commercial purposes. These concerns are understandable and justified.

The role of the libraries not to downplay these concerns, but on the contrary to be a *safe haven*, a place where everything that is private is respected and protected.

-.-