Paper assesses whether there is a set of agreed upon principles or disagreement over what constitutes “ethical AI;” the paper's analysis finds a lot of overlap among the sets of principles analyzed (what sets of principles are these?)

The paper then summarizes the “techno-scientific” (I assume this to mean a combination of technological and scientific) branches of AI. Image/content generation is not in the figure that does this, but natural language generation is shown as a form of natural language processing in machine learning. However, the AI is not “intelligent” or even “thinking” (i. e. The Weak AI argument as it's called in philosophy). As such, AI is conceptualized as a resource of “interactive, autonomous and self-learning agency” that can deal with tasks that would usually need a human’s intelligence to do.

Each set of proposed principles are recent (published within the last 3 years), relevant (directly relevant to AI and its impact on society) and reputable (published by organizations that are at least on the national level)

There is some coherence and overlap between 6 sets of principles (what does it mean by 6 sets if there are only 5 principles proposed and discussed in the paper?); this is mostly seen in the overlap between 4 core principles used in bioethics: beneficence, non-maleficence, autonomy and justice. These 4 principles adapt well to ethical challenges posed by AI, but they are not exhaustive and the names of the principles are often used to refer to different things. Therefore, the paper argues that the additional principle of explicability (incorporates both intelligibility and accountability) is needed. The paper also warns that the overlap between the different sets of principles demands caution.

5 principles:

* “Beneficence: Promoting Well-Being, Preserving Dignity, and Sustaining the Planet”
  + Creating AI technology that is beneficial to humanity
* “Non-Maleficence: Privacy, Security and ‘Capability Caution’”
  + Create beneficent AI but also be wary of the consequences of the overuse or misuse of AI
    - Prevention of infringements on personal privacy is of particular concern and is included as a principle in 5 of the 6 sets (do “documents” and “sets” refer to the same thing here? As in the 6 sets of principles even though there's 5 principles?)
* “Autonomy: The Power to Decide (to Decide)”
  + Balance our decision-making power and that which we give to AI
* “Justice: Promoting Prosperity, Preserving Solidarity, Avoiding Unfairness”
  + AI development should seek to eliminate any and all discrimination
  + Need both “shared benefit” and “shared prosperity” from AI
* Explicability: Enabling the Other Principles through Intelligibility and Accountability”
  + Crucial gap in AI ethics: how does AI work (intelligibility) and who is responsible for how it works (accountability)?
  + Complements other 4 principles

Former 4 principles are traditional bioethics principles and the latter (explicability) is a new enabling principle for AI.

The authors of the article believe that every company, government agency and academic institution designing, developing or deploying AI has an obligation to do as such in line with an ethical framework like the one proposed in the article since ethics is not the “preserve” of a single continent or culture. They also write that laws, rules, standards and best practices to regulate or control AI would also benefit from working closely with a unified ethical framework.

According to the table in the paper, all the AI programs analyzed have both nonmaleficence and explicability, which seem like the most important ethical principles for this research work.

DOI to paper: [https://doi.org/10.1162/99608f92.8cd550d1](http://doi.org/10.1162/99608f92.8cd550d1)