The five laws of ethical digital human design

In 1942, Isaac Asimov introduced us to the "three laws of robotics" – designed as a series of safeguards to prevent AI from harming humankind.

Today, our understanding of the effects of AI is more nuanced; we understand that while Asimov's rules prevent artificial intelligence from physically hurting us, it doesn't stop real people from designing AI to mislead, to spread fake news and to cause other types of harm and distress.

The Five Laws of Ethical Digital Human Design form a foundation for ensuring AI-powered digital humans are able to live their best virtual lives, for our benefit.

And like any foundation, we want to build on it – with your help.

A **principle-led** approach to building a world with **digital humans**.

Digital humans are on the brink of becoming a part of our everyday lives – in how we live, how we work and how we manage our health. And while they aren't "real", their implications on people very much are.

We become emotionally connected with our digital human coaches, personal assistants, health experts and advisors because of the human-like qualities and conversations we share. So, we need to protect **real people** from those who would use the trust people build with AI to cause harm.

These five laws are written to:
- Promote the positive use of AI digital humans; and
- To bring out the best side of humanity from day one with a set of community-created standards for designing, employing and interacting with digital humans.

Although there is currently no legislation around adhering to these principles, those who don't hold themselves to public scrutiny...

If they don't, why not?

Digital humans shall never pretend to be what they're not.

Digital humans build genuine relationships with people based on trust, so they shouldn't be used to deceive.

As we head into an era where it's difficult to distinguish real from created, it should be made transparent when a user is talking to an AI digital human.

In addition, digital humans shouldn't be created in the likeness of other people without their permission.

Digital humans should always be designed to help not harm.

Digital humans should exist to improve lives only, and not cause emotional or physical harm to humans in any instance. This includes restrictions on forcing digital humans to perform tasks that reduce the quality of human life—such as being used in illicit or illegal activities, or to voice hate speech or fake news.

Digital humans must respect the privacy of those they interact with.

Creators must respect people's right to privacy. Digital humans should not capture nor store personal data without permission, nor use more personal data than is necessary. They must be upfront and clear about what data is used and maintain the individual's rights over their data.

Digital humans should be designed to **uphold respectful behavior.**

Digital humans should be designed to be respectful, but also empowered to expect respect in return. How might digital human experiences be designed so they don't enable disrespectful behavior—e.g. disconnecting automatically after a discriminatory attack?

We should be careful not to create "subservient" digital humans who reinforce a habit of talking down to people who are of service to us—regardless of whether they're real or not.

Ethical design should aim to create digital humans to help us be better people. Digital humans who work in mental health and education, for instance. Think of those people in everyday life who have inspired and supported you to be a little bit better. How can digital human design (from personality design to conversational design) replicate similar supportiveness?

Co-design of digital humans should be adopted to avoid bias and promote diversity and inclusive representation.

Creators must strive to include input from a range of people and data to ensure we design digital humans that represent the diversity of society as a whole.

Creators should also strive to solve accessibility challenges, including making the technology available regardless of disabilities or socio-economic standing. This should involve thinking about what it means to serve the underserved—accessibility and literacy are just some of the examples that we should bring into this consideration.

As far as possible, creators should work alongide people and groups who understand the challenges the digital human is trying to solve for, and involve diverse views and inputs so the end result is a digital human population that caters for all.
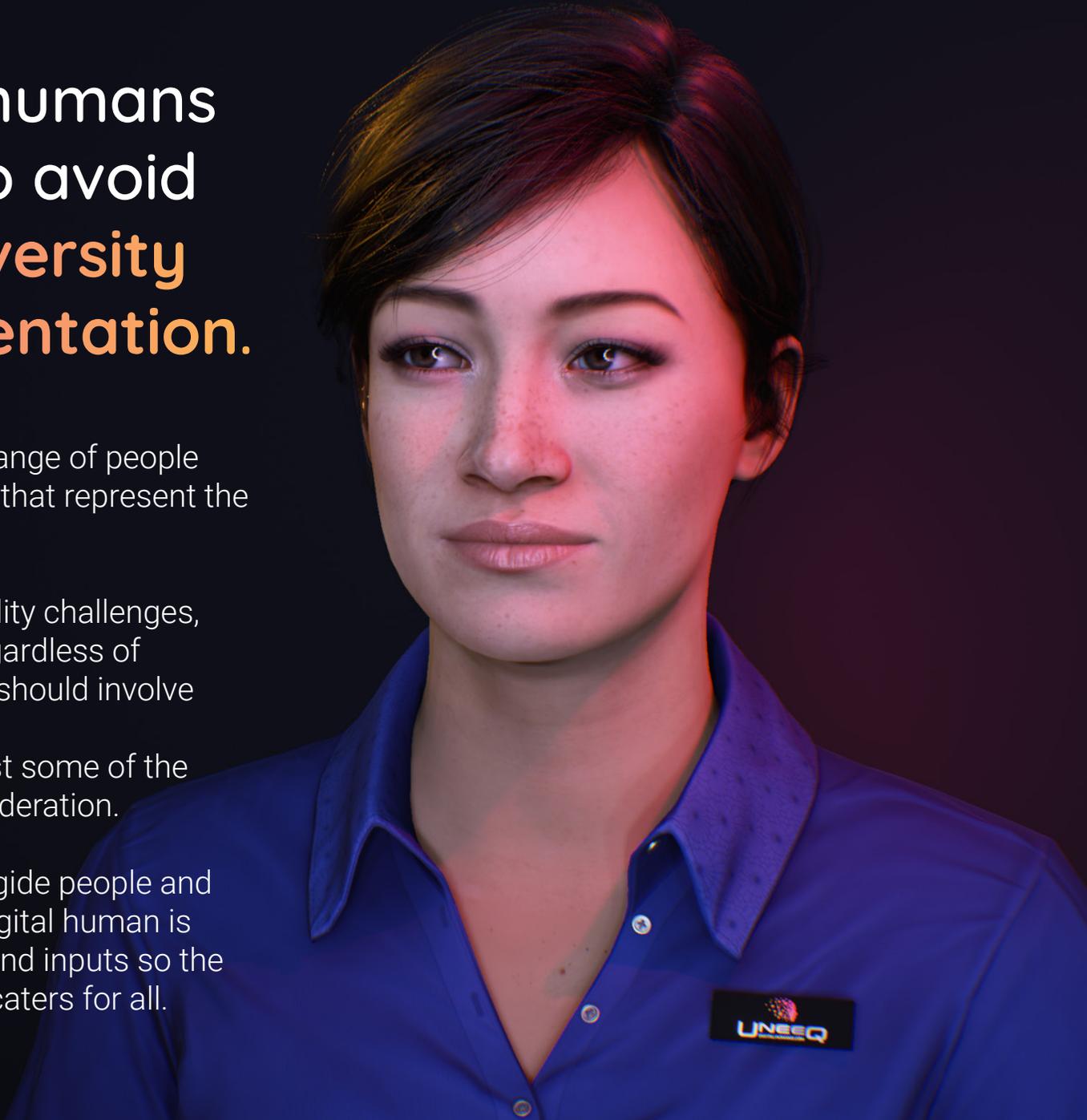
HAVE YOUR SAY

The conversation has started. Contribute to the Laws of Ethical Digital Human Design so as a community we can make it better and fairer for all people – digital and otherwise.

FIND OUT
MORE HERE
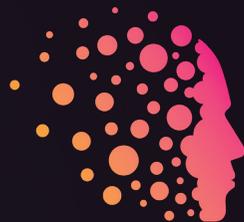
UNEEQ
DIGITALHUMANS.COM

Learn more about digital humans
at digitalhumans.com