

On the digital intelligence across modalities: Cheers and fears

Mohamed Elfeki,
Microsoft Research, USA

Abstract

Humans observe and comprehend the world through their senses: sight, hearing, touch, smell, and taste. Most primate newborns -including human babies- are born with all the five senses in which they use in a complementary fashion to understand the events and entities surrounding them. However, modern Artificial Intelligence (AI) decided not to fully assimilate human development. That's simply because AI models are developed in the cyberspace which contains different types of entities to recognize. As opposed to neonates born in hospitals surrounded by doctors and nurses, the developed AI models are likely to be born on the internet surrounded by millions of Instagram images and billions of Google words. Hence, AI scientists are focused primarily on teaching the digital models its surrounding environment, which happens to focus on two of the five signals that human use: sight (images/videos) and hearing (audio/text). This work discusses developing AI models in digital environments and analyze the mechanisms and implications of using those two senses to learn. The technology and progress of AI has been unpredictable, unprecedented, and full of surprises. There's a lot to applaud but also quite some to fear. Ultimately, the models developed will be used by humans and they're only amplifying our deepest desires, the best as much as the worst. We can only hope the next few years will demonstrate the most noble of the human race

Received: March 12, 2022; **Accepted:** March 21, 2022; **Published:** March 29, 2022

Biography

Mohamed Elfeki is an Applied Science Lead at Microsoft Research, USA. Prior, he experienced different job titles in several top research labs including Facebook Artificial Intelligence in Paris, Siemens Research in Princeton, and Amazon Research. He obtained his Ph.D. from University of Central Florida in 2020 working on a wide variety of

computer vision and machine learning topics including image/video synthesis, multi-object tracking, dynamic modeling, video summarization, zero-shot learning, metric learning, meta- learning, and domain adaptation. Currently, he's leading a science team developing state-of-the-art solutions for product-oriented researcher.