

Given **two observations**:

1) The bottleneck in AI is **access to data**

2) Most of **today's data** is **personal** one way or the other

What is the role of **interoperability** (data APIs), multi-homing, and **data portability** in competition law and regulation moving forward?

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough:		3 years	18 years



A) The average elapsed time between key algorithm proposals and corresponding advances was about 18 years, whereas the average elapsed time between key dataset availabilities and corresponding advances was less than 3 years. [from [Alexander Wissner-Gross](#)] B) Most of today's data contains information about humans one way or the other and anonymized behavioral data can often be re-identified [from [Sandy Smolan](#)]