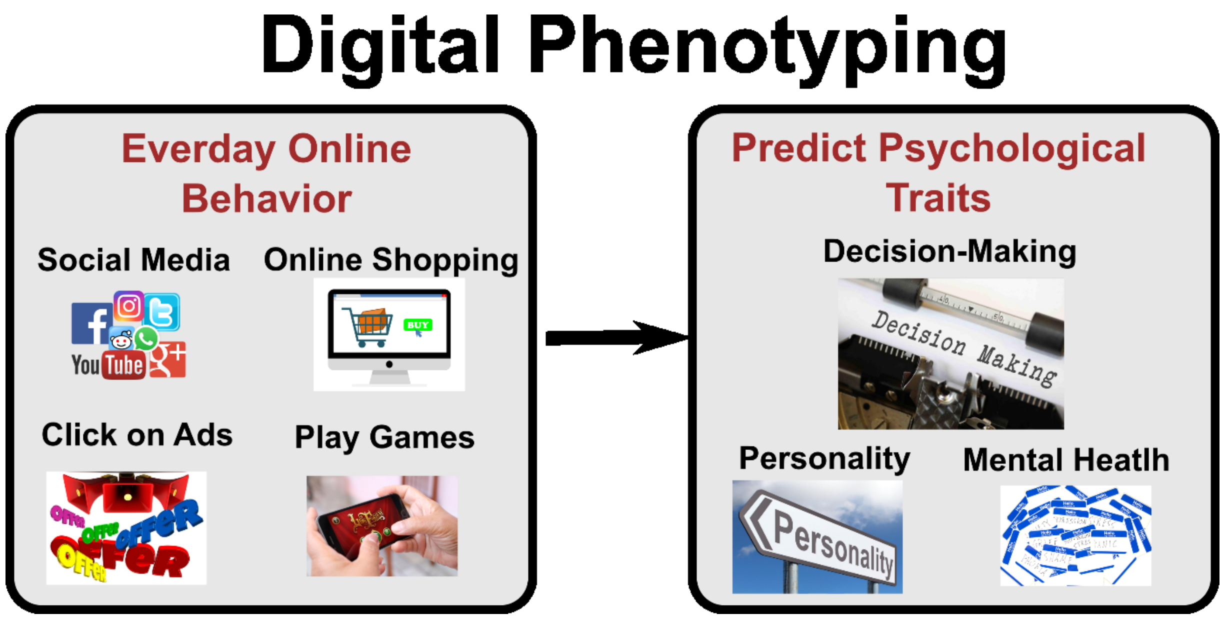
Data Mining the Mind: A Research Statement

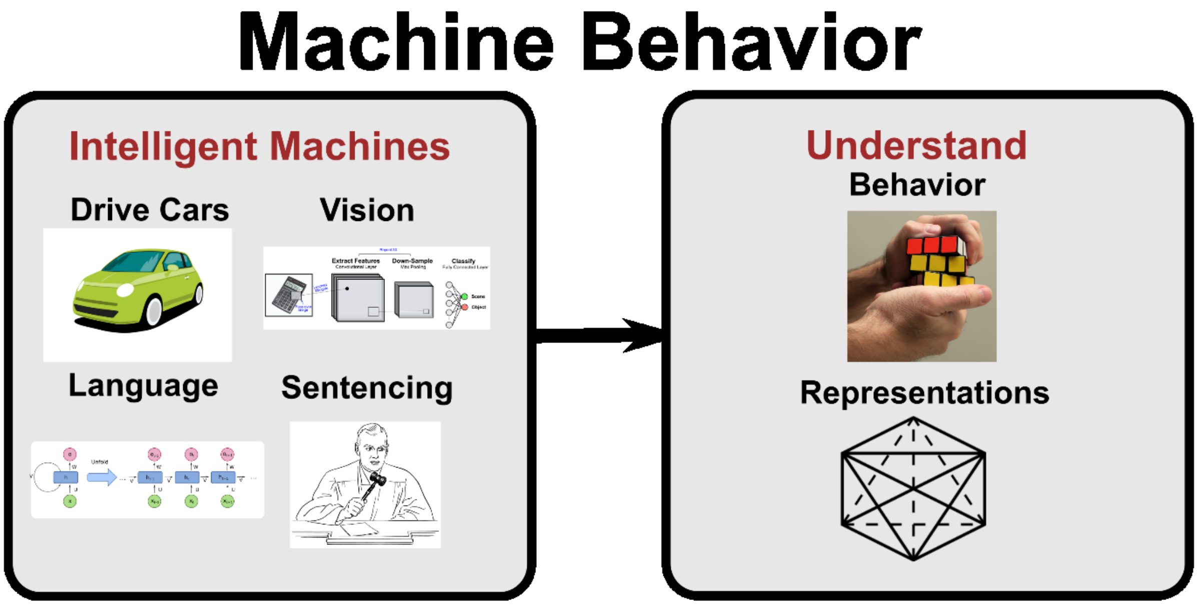
Robert Thorstad

The past and coming decade have seen enormous developments in technology. A majority of US adults now use social media at least once a day (Perrin & Anderson, 2019), and terabytes of data about human behavior are recorded every day, as when people shop for products, search for information, or click on ads. In the coming decade, people will increasingly create intelligent machines that perform cognitive tasks typically performed by humans. There is progress being made on machines which drive cars, understand language, process images, and deliver criminal sentences.

These developments are not just technological, but fundamentally raise two psychological questions.



<u>1) Digital Phenotyping: How Much Does Our Everyday Behavior Reveal About Our Psychology?</u> The average twitter user has posted more than a thousand tweets online, each of which is a seemingly innocuous snapshot of everyday behavior. When this information is aggregated, I ask whether beyond what is explicitly said, this everyday behavior can reveal deep facts about our psychology. Thus far, I've shown that social media posts are predictive of people's real-world intertemporal decision-making (Thorstad & Wolff, 2018) and of whether they have a mental illness (Thorstad & Wolff, in press). The long-term goal is to answer questions like: what traits can be inferred from people's online behavior? Is there any psychological trait a digital phenotyping algorithm can't infer, given enough data? Do these algorithms fundamentally work like people (who can make some of the same inferences based on snap judgments known as thin slices), or are they fundamentally different than people?



2) Machine Behavior: How do Intelligent Machines Behave and Why? Increasingly, machines are achieving human-level accuracy at tasks like vision and language, with improvements in driving and sentencing algorithms on the horizon. We know a lot about how accurate these algorithms are, but these tasks necessitate a much more sophisticated behavioral picture than accuracy. When a machine learns to see or read, what has it learned? Is the algorithm biased, and if so, how? When a self-driving car decides between hitting a pedestrian and colliding with another car, how does it make this decision? These questions necessitate studying autonomous machines as if they were human or animal agents: a science of machine behavior (Rahwan et al, 2019). Thus far, I've shown that computer vision models trained only to distinguish scenes from objects learn previously undiscovered global shape cues – a scene's horizontal horizon – that are also used in scene-selective regions of the human brain (Cheng, Thorstad, & Dilks, in prep).

I answer these questions by building large-scale, functioning models. My approach is to build machine-learning models that can accomplish a task given large, varied, real-world sets of stimuli (often hundreds of thousands or millions), and do so with high accuracy. Having built these models, I then ask, either: What can the model infer about a person's psychology? Or, how does the model perform its task, what does it know, and what can this teach us about human psychology? This typically involves computational modeling, but I also use human experiments to compare humans with the models. I sometimes also compare the model's representations to those used by the human brain.