BEYOND THE HYPE

Innovation predictions in the era of machine learning

July 2022

AUTHORS

Colin Ho, Ph.D.

Jiongming Mu

Rich Timpone, Ph.D.

Camille Souchal

IPSOS VIEWS

GAME CHANGERS

Ipsos

The use of Artificial Intelligence (AI) has been booming in recent years. Programs such as voice and facial recognition are embedded in cell phones, televisions, cars, and other consumer products (e.g., Amazon Alexa). Algorithms power robots to perform tasks and help us choose gifts and find places of interest. A PWC study estimates that AI will boost global economic output by at least $15 trillion by 2030, an amount greater than the current output of China and India combined.[1]

Against this backdrop, we are also starting to see how AI will similarly revolutionize market research, leading to faster, cheaper, and better outcomes. However, AI has many different definitions which has led to a lot of confusion. This confusion is exacerbated by those who feed into the hype and benefit from exploiting the ambiguity.[2]

One reason why it has been difficult to achieve a good understanding of AI is that the term covers an extremely wide range of technologies and applications. A quarter of a century ago, we identified this diversity as an issue in understanding AI as, at the time, there were different domains including: 1) the study and application of human knowledge and intelligence, 2) the study and development of machine intelligence, and 3) the application of intelligent algorithms to solve practical problems.[3] Since then, this last category has exploded through diverse methods, primarily leveraging Machine Learning methods which are now impacting our lives everywhere.

We will spend more time on defining and discussing how to evaluate the accuracy of these methods and their ethical implications in other POVs. In this paper we will focus on the development of AI and Machine Learning (AI/ML) to build models to accurately predict a new innovation's performance using consumers' language. While the nature and quality of the algorithms are important, traditional market research questions of data representativeness, measurement, and relevance represent critical areas which identify why the blending of human intelligence with artificial intelligence is so important for how we leverage AI/ML to predict the potential of new innovations.

WHAT IS PREDICTIVE ANALYTICS AND HOW DOES IT RELATE TO AI/ML?

Predictive analytics is the use of statistics, machine learning and algorithms to make predictions about the likelihood of **future outcomes** based on historical and/or current data. In addition to predicting outcomes, the term predictive analytics is sometimes also used to include models to understand the **drivers** of a particular outcome so marketers can better understand which levers to pull to create the future they want. There are many different **data** predictive models can be built from (e.g., time series data, social data, survey data). Determining how a predictive model will be used (e.g., what outcome are we trying to predict? Do we need to understand the drivers?) and what data sources are available to build the model are two questions a user of predictive analytics should always ask themselves. The answers to these questions will help guide the direction forward.

While some distinguish predictive analytics by focusing on the new areas of 'technology that learns from experience (data)',[4] the reality is that prediction methods have been around for quite a while. Market research approaches that have prediction at their core, however, have not been widely recognized as using predictive analytics because of differences in terminology between market research and AI (e.g., forecasting vs predictive analytics) even when they are using AI/ML techniques.[5] To be fair, the *components* that make up predictive analytics have evolved dramatically. Specifically, the data sources available for predictive analytics have grown considerably (e.g., sales data, social data, behavioral data, video data), as have the analytic algorithms available (e.g., machine learning and deep learning algorithms which do differ in some goals from statistical methods).[6] Vast improvements in computing power have also made the large-scale applications of predictive modeling to big data possible. These developments are all being used in market research and are fundamental to many Ipsos services (e.g., predicting trends from social data; using video tags to predict the performance of advertisements).

Rather than providing a theoretical overview around the range of predictive analytics, this paper focuses on one specific domain - the use of Machine Learning to predict the success of new product innovations. By using concrete examples without unnecessary jargon, we will provide more clarity around the subject. As we will see, the nature of the data for training predictive analytic models is critical and this will be illustrated in how Ipsos' predictive model of innovation success directly addresses the practical challenges involved.

TRAINING AN AI/ML MODEL TO PREDICT NEW PRODUCT INNOVATION SUCCESS

THE IMPORTANCE OF TRAINING DATA TO MACHINE LEARNING

When we want to teach a child to recognize a phone, we show the child what a phone looks like. Better still, we show the child examples of phones and non-phones, so the child learns to recognize features unique to phones (e.g., rectangular glass screen, camera on the back). By so doing, the child learns to identify all phones, not just the examples provided (see Figure 1).

Training an AI/ML model to recognize a successful new product innovation is conceptually similar. We need examples of successful and unsuccessful innovations so an AI/ML model can learn to recognize the features that predict successful innovations. With enough examples, a trained model can predict the success of innovations not previously seen. Training is the foundation of all predictive models, and the quality and appropriateness of the training data used will determine how good your model will be.

In the context of predicting new product innovation success, the training examples most people think of are innovation concepts tested previously. This likely comes from what people have seen in popular media. For example, we teach an AI to play chess by giving it examples of chess moves and the associated outcomes. If we want to teach an AI to recognize human faces, we provide the AI model with examples of faces. While this works in many areas, this does not work as well when it comes to building predictive models for innovations. Unlike other areas where what you are trying to predict does not change (e.g., the rules and the pieces used in chess stay the same), or where change occurs at such a pace that the updates to training can be accurate, predicting innovations, by definition, means we are trying to predict things that ***don't exist today***.

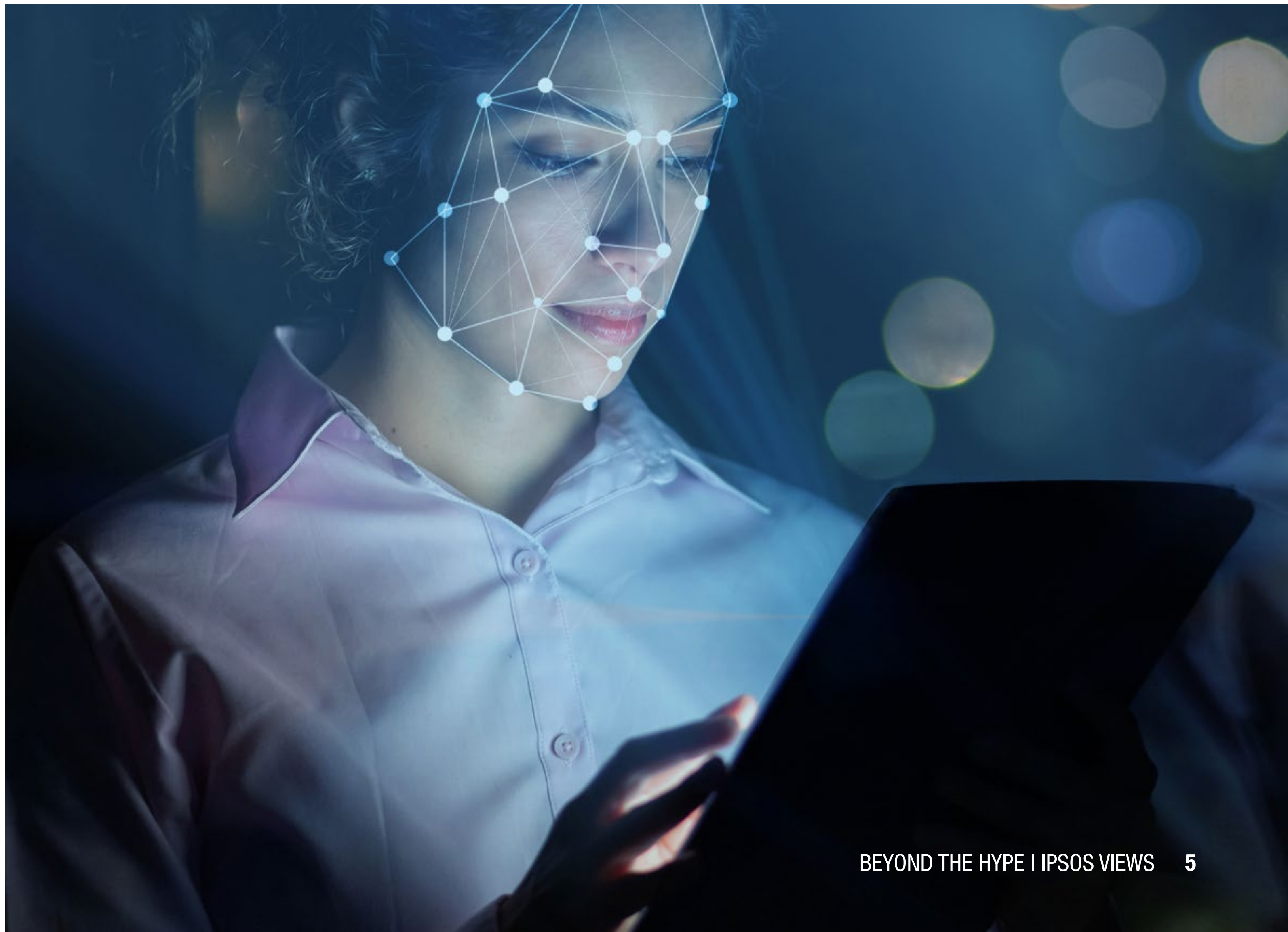
As we will show, this does not mean that predictive analytics can't be applied but rather that the choice of training data is critical, and in this case many available sources do not address the dynamic environment. Given the critical role training examples play in building predictive models, we will elaborate on this point further.

> If we want to teach an AI to recognize human faces, we provide the AI model with examples of faces. While this works in many areas, this does not work as well when it comes to building predictive models for innovations. ”

Figure 1 An illustration of how learning from the past can be helpful for predicting a future not too different from the past

CHILD SHOWN EXAMPLES OF WHAT A PHONE IS ...

ACCURATELY IDENTIFIES A NEW PHONE



THE PAST MAY NOT PREDICT THE FUTURE

In our child learning example, the implicit assumption is that phones today will look like phones tomorrow. Consider a scenario when new phones look completely different (e.g., a wrist band voice-activated phone with no number pad, like Apple Watch). In this scenario, the examples used to identify a phone from the past would not be useful for identifying ones that look completely different (see Figure 2).

The relevance of training examples to the future is critical when it comes to training AI/ML models to predict product innovation success. Let's imagine that we train an AI/ML model to predict the success of new personal cleaning products by providing it with examples of personal cleaning concepts tested previously. To make it even more concrete, let's further imagine that the concept examples used for training include concepts of the successful Ivory brand of bar soap launched in the late 1800s.

Part of Ivory bar soap's appeal then was that it floats, a benefit that would be desirable when taking a bath, so you don't need to grope around on the bottom of the bath to find it. This benefit, however, is less relevant in today's world where showers are more common.

Put simply, when it comes to product innovations, product features of **innovation concepts from the past may not predict successful innovation concepts of the future**. Products and benefits that were successful at one point in time may not be relevant down the road. Over time, new products introduced into the market change people's expectations of which benefits products should offer. Think of how phones have changed in the last decade and, along with that, our expectations of what a phone should be able to do. When the future can change from today, using past or even current product examples to train a predictive model is generally not a good idea.

Figure 2 An illustration of how learning from the past may not be helpful for predicting a future that is quite different

CHILD SHOWN EXAMPLES OF WHAT A PHONE IS ...

UNABLE TO IDENTIFY A FUTURISTIC PHONE THAT IS DIFFERENT

A RAPIDLY CHANGING WORLD

It's not just products that change and lead to different expectations. The world itself changes and leads consumers to prioritize and value different benefits or seek new benefits. We only need to look at climate change, the current pandemic, and technology, to see how changes in our world can drastically change what consumers value.

In the past year or so, there has been a growing interest in sustainable products. Given the recency of this development, a concept-based model using concepts tested in the last five years would have very few examples to teach a predictive model to recognize whether sustainable products will be successful or not. The pandemic, another large change in our world, has led to a greater prioritization of health – especially when it comes to boosting one's immunity. Even if immunity was a desired benefit five years ago, the pandemic has increased the desirability of immunity considerably. A predictive model trained on concepts tested in the last five years may, therefore, underestimate the importance of immunity benefits today. Finally, consider how rapidly consumer packaged goods (CPG) are evolving in the digital world. We now have, for example, vitamins that can be personalized by answering a series of questions online (e.g., Care/Of), and children's toothbrushes that offer an augmented reality toothbrushing experience (e.g., Colgate MAGIK). A CPG concept database accumulated in the past five years would not have any examples of personalized or augmented reality products using digital technology.

The impact of a rapidly changing world on AI models was seen very practically in the early period of Covid when existing AI/ML models from health to shopping behavior failed because people's behavior changed drastically. Models that had been trained pre-pandemic were not able to predict well in a world that had changed.[7]

The general point is that we cannot predict how the world or products will be in the future from the concepts themselves. Using concepts from the past few years or even today as training examples may lead to the complete omission of new benefits (e.g., personalization of a product) or result in overestimation or underestimation of the importance of existing benefits (e.g., immunity) that change with the context. A concept-based predictive model may predict well for a year or two but will quickly become outdated with rapid market changes, and **where it does perform well is for products that look like others already in the market and not truly innovative ones**. The digitization of our world, the pandemic, and climate change have all illustrated how quickly the world, and our needs, can change.

CONSUMERS RESPOND TO THE TOTAL PROPOSITION

With new product innovations, consumers respond to the total proposition, not individual components described within a concept (e.g., insight, benefits, reason to believe, price). A concept-based model based on independent features and benefits in the descriptions, therefore, may not accurately capture the appeal of a new product innovation. The various components of a new product concept often interact and have an interactive effect on how consumers respond. Consumers' reaction to a new product goes beyond a simple sum of its parts.

As an example, a concept-based model may determine that avocados and hot chili sauces, separately, lead to higher food concept scores when they are included in the description. If both are present in a concept, a simple model would predict a robust positive response. In truth, if a concept featured an avocado hot chili smoothie, consumers would most likely react negatively! Consumers react to the entire proposition and not just individual elements.

USING PREDICTIVE ANALYTICS PRUDENTLY IN INNOVATION TESTING

ACCURACY: EVALUATING THE EFFECTIVENESS OF THE LEARNING

When a child first learns to identify a phone, the child might make mistakes when they encounter phones not seen previously. The same applies to AI/ML models. Predictive models are never 100% accurate. There is a trade-off when we use a predictive model (i.e., decreased accuracy but increased speed and cost efficiencies for predicting future possible outcomes). Assessing the level of accuracy of an AI/ML predictive model is critical for making an informed judgment on the suitability of the model for the application in mind. Because predictive models are imperfect, our recommendation is to use innovation testing predictive models on early-stage concepts when the goal is to screen many concepts for more rigorous testing later.

AI/ML models are validated on a set of examples that were not used to train the model but where the outcomes we are trying to predict are already known. This means evaluating the AI/ML model on product innovations it has not seen before, but where we already know whether the innovations were successful or not. If the model's predictions match perfectly with how the examples performed (i.e., successful, or not), then accuracy is 100%. In practice, accuracy for most product innovation predictive models falls between 60-80%. Determining what is an acceptable level of accuracy is a judgment call, usually done by weighing the risk involved in making a wrong decision against the benefits of using a predictive model.

It's worth noting that the definition of what is considered successful or unsuccessful can be subjective (e.g., relative to expectations, to a market average, compared to competition, etc.). Whatever the metric, a user should always know the accuracy of a model, how it is computed and the metric it is based on.

Assessing the level of accuracy of an AI/ML predictive model is critical for making an informed judgment on the suitability of the model for the application in mind. ”

NEW PRODUCT PREDICTION IS SPECIFIC

Returning to our child example, a child that has been taught to recognize a phone would not be expected to recognize non-phone items (e.g., coffee machine, refrigerator). Much of what we learn tends to be specific, and the advances in AI/ML already mentioned are in weak or narrow AI where algorithms solve single tasks. That is why the AI for image recognition, chess, and speech recognition are developed using different training examples and validated separately, and this is true for pairs of similar tasks as well (e.g., an AI learning how to produce images like Rembrandt vs Monet would be separately trained even if they leverage the same algorithms). The narrowness of training learning models also applies when predicting innovation success from relevant data.

The tenet that you predict only what you trained the model on has an important implication in the application of AI/ML models in product research. If we build an AI/ML model to predict the success of beverages, we should not use the model to predict the success of household cleaners. If we build an AI/ML model to predict the success of beverages in the US, we should not use the model to predict beverage success in China. This is because the predictors of success for each product category are different, and what is important to consumers in one country may not be important to consumers in another country.

Users of AI/ML predictive models need to know the examples a model was trained on and whether the predictors in the model were built for the specific application they will be using it for. Knowing this will help the user assess how broadly the model can be applied. A trained model needs to be retrained if its purpose changes or if it's used in a different environment. You cannot take the same model that was trained to fulfill one task and then expect it to work well in a different context. AI/ML isn't miraculously good at everything.

While more general AI/ML models that can distinguish category and geographic differences could be created, the amount of relevant data to address such heterogeneity is beyond what is usually available. Cutting corners with insufficient or less relevant data, as described with concept data, will create poor accuracy for future prediction due to differences.

> The AI for image recognition, chess, and speech recognition are developed using different training examples and validated separately, and this is true for pairs of similar tasks as well.

IPSOS' PREDICTIVE MODEL FOR INNOVATIONS

Given our previous discussion on concept-based models, it should not be a surprise that we do not use concept content as the foundation for prediction. Instead, we rely on **consumers' verbatim responses** to concepts to predict performance. Essentially, consumers are shown a new product concept, and then asked for their **top-of-mind reaction** to the concept via an open-ended question. It is this immediate reaction to a concept that is used in our predictive model.

We use consumer reactions to train our model and predict the success of new products because we believe that consumers' verbatim responses are robust and stable **predictors of innovation success**. How consumers react to new product ideas is unlikely to change much over time. A decade from now, we expect consumers to continue expressing positivity when they like a new product idea, negativity when they dislike a new product idea, or indifference when they have no interest. Using consumers' verbatims allows us to build AI/ML predictive models leveraging past data (consumers' verbatim reactions). This comes with a trade-off: we need to collect, at minimum, consumers' verbatim reactions to predict the future performance of new product ideas.

TRAINING DATA

As of February 2022, we have accumulated about **four million** consumers' responses from testing new product concepts across 60+ countries and seven mega-categories (human food, beverages, health care, homecare, personal care, beauty care, and pet care). This database provides the data to train our predictive models, and ensure they are robust across a wide range of concepts and responses. As this database is automatically populated with new studies continuously, our database also allows for regular model updates.

The two sets of items we use from our database to build our predictive models are 1) consumers' top-of-mind verbatim responses to new product concepts and 2) whether they chose the new product or their existing solution on three key metrics validated to actual new product launches – Relevance, Expensiveness and Differentiation (RED).[8] Essentially, consumers' immediate reactions to new product concepts are used to predict if they would choose a new product or choose to stay with their existing solution.

Consumers' top-of-mind responses are in the form of raw verbatims collected from a single open-ended question asked immediately after the concepts have been viewed. As examples, verbatim responses to food concepts may include positive responses like "yummy" and "makes me salivate". They may also include negative responses like "gross" and "not for me". It is these verbatims that are used to predict whether people will choose their new product or their existing solution on the three key metrics. **How** people respond will always indicate their positivity or negativity towards new products, **so** the examples use to train the model **will** remain relevant to the future and ensure the predictive model can anticipate future success and does not become outdated.

While we understand the preference for predictive models that require no consumer input, for the reasons discussed earlier, a minimum level of consumer input is required. With only one open-ended question, new product innovations can still be screened expeditiously and cost effectively. While acknowledging the limitations already noted, we plan on continuing our evaluation of methods to see if and how the content of concepts can add incremental prediction at a later stage, but this would be above and beyond the prediction using consumers' verbatim response which is foundational.

> How people respond will always indicate their positivity or negativity towards new products.

ACCURACY AND SPECIFICITY

It is against the three RED metrics that our predictive models are validated. That is, we compare what our model predicts against the actual choices respondents made in previous concept tests. As an example, if 100 respondents selected the new product when choosing between the new product and their existing solution, and our model *predicted* that 75 respondents selected the new product and 25 selected the existing solution, then the accuracy would be 75%. Based on the models we have built to date, the accuracies of predicting the RED metrics range from 70-75%. As the goal of early-stage screening is to ensure a winner progresses to the next stage for more rigorous testing, this level of accuracy provides an acceptable trade-off for greater agility and cost efficiency.

It should be noted that accuracy levels apply only when the model is applied to new concepts that fall in the same category/country as those used to train the model. For example, the accuracy of a US food model refers to the model's ability to predict new US food concepts only. If there is a need to apply predictive modeling more broadly, for example, to countries other than the US, it may be better to build a global food model that allows for broader application. A global food model may have slightly lower accuracy than a US-specific one but will allow for broader application across countries.

A FINAL NOTE

We have sought to demystify AI/ML and explain in simple language how such models can be developed and used for predicting new product innovation success. We have made it clear that the promise of predictive analytics depends heavily on the relevancy of the training data to the future. To use these models judiciously, a good understanding of the trade-off involved (e.g., accuracy vs cost) and of the limits of a model due to the narrowness based on the data available is also needed.

AI/ML models for predicting innovation success can make innovation testing faster and more cost efficient, cutting down project timelines from weeks and months to hours and days. But like any research or analytic approach, a strong understanding is needed to ensure proper usage and application and so we focus on augmented intelligence in innovation testing, combining human and machine learning for the greatest value.

The era of machine learning is here indeed but let's ensure we adapt how it is used in market research. If you have ever invested your money and read through the prospectus of investment companies, you may have noted the caveat "past performance is not indicative of future results". It's the same when you are trying to predict the success of future products using the performance of old products. Building a predictive model to determine the success of things that don't exist today is very different than building an AI/ML model to play chess, drive a car, or suggest movies you might like.

REFERENCES

1. Horgan, J. (2020). Will Artificial Intelligence Ever Live Up to Its Hype? Replication problems plague the field of AI, and the goal of general intelligence remains as elusive as ever. Scientific American online https://www.scientificamerican.com/article/will-artificial-intelligence-ever-live-up-to-its-hype/

2. Olson, P. (2019). "Nearly Half of All 'AI Startups' Are Cashing in on Hype." Forbes https://www.forbes.com/sites/parmyolson/2019/03/04/nearly-half-of-all-ai-startups-are-cashing--in-on-hype/#5635ecf1d022

3. Taber, C.S. and Timpone, R.J. (1996). Computational Modeling. Sage University Paper Series on Quantitative Applications in the Social Science, series no. 07-113. Thousand Oaks, CA: Sage

4. Siegel, E. (2013). Predictive Analytics. Hoboken, NJ: John Wiley & Sons.

5. Venugopal, V., & Baets, W. (1994). Neural Networks and Statistical Techniques in Marketing Research: A Conceptual Comparison. Marketing Intelligence and Planning, 12 (7), pp 30-38

6. Agrawal, A, Gans, J., and Goldfarb, A. (2018). Prediction Machines. Boston, MA: Harvard Business Review Press

7. Heaven, WD. (2020). "Our weird behavior during the pandemic is messing with AI models," MIT Technology Review https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop

8. Ho, C., Mu, J., Brown, A. (2019). Speed Dating with Innovations: What's Your Innovation Pick-Up Line? Ipsos POV. https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-06/innovation-speed-dating-ipsos-views.pdf

BEYOND THE HYPE

Innovation predictions in the era of machine learning

AUTHORS

Colin Ho, Ph.D.

Chief Research Officer, Innovation and Market Strategy and Understanding, Ipsos

Jiongming Mu

Global Innovation Solutions Lead, Ipsos

Rich Timpone, Ph.D.

Managing Director, Global Science Organization, Ipsos

Camille Souchal VP,

Innovation Technologist, Ipsos

The **Ipsos Views** white papers are produced by the **Ipsos Knowledge Centre.**

www.ipsos.com

@Ipsos

GAME CHANGERS