Building Breakthrough Products Using AI Research
Nick Ghotbi and Ryen White, Microsoft Cortana

Artificial intelligence (AI) plays an increasingly important role in our lives. More and more of the applications and services we use each day are enhanced by AI technology. One example is Portrait Mode on the Apple iPhone, where AI is used to intelligently blur photo backgrounds in real time. Integrating AI research outcomes into products is a challenging endeavor that requires close collaboration between research and engineering teams.[1] We address the challenge of taking existing breakthrough research—sourced internally within an organization, or externally, e.g. in academia—in a new area where there is clear market opportunity for building a product around it. We ground this with examples of the work our team did on imbuing Cortana, Microsoft's digital assistant, with new AI capabilities centered on identifying "commitments" (promises) made in outgoing emails. We offer the lessons learned on our journey and the steps you can take to ship products and services augmented by groundbreaking AI research.

1. Focus on a unique user problem and find the right research
Before looking for any type of AI technology or research to integrate into an existing product or to build a new product around, it's important to start with the user problem that you want to address. Starting your search for the AI with that problem in mind will help scope your search and find the right technology more quickly. As an example, in Cortana, we were looking to innovate around augmenting human memory and helping people be more productive. As part of this, we decided to specifically focus on the user problem and anxiety around forgetting to do things promised to others (such as telling your manager "I will send the updated budget by end of week" or promising your partner "I'll make the restaurant reservation for this weekend"). Once this niche problem was validated through market and user research, discussions with scientists in our company's research division, Microsoft Research led us to existing AI research that could be applied to solving the problem through an experience in Cortana.

Extracting commitments from communications isn't a new concept. Some of the early work to demonstrate the feasibility of using machine learning for this task was done at Microsoft over ten years before we started work on delivering it as part of the Cortana product. The commitments feature therefore provides a canonical example of the challenges in productizing existing AI research. Although the research was already done, our team still needed to define the appropriate product market fit and spend considerable time revisiting prior assumptions and adapting the technology to product needs. Now that the experience has shipped, it is helping many Cortana users keep their promises each day. Microsoft CEO Satya Nadella even mentions the commitments experience in his recent book, Hit Refresh, as a Cortana feature, and example of AI, that he regularly relies on.

Not all companies have access to their own research division. Luckily, you can find existing AI research in a variety of places, including coverage of technical advances in the popular press and scientific journals and conferences, which are often archived and searchable in digital libraries. Scientific conferences such as AAAI and NIPS are an effective way to learn about breaking AI research, although the content can be highly technical. Industry events including The AI Summit provide insights and strategies for AI practitioners. Technology industry news sites such as TechCrunch and Ars Technica describe new AI developments in a broadly accessible way. Specialized search engines such as Google Scholar or Microsoft Academic help find scholarly content if you know what you are looking for.

---

[1] See the "hybrid research model" [1] employed at Google for a good example of this approach.

2. Build a team of experts who are focused on the user problem

To productize existing AI research, you need to build a team of people aligned along many dimensions, especially the user problem being addressed and the product goals. The team needs to be multi-disciplinary, consisting of team members who can understand the research that was done and make modifications to it as needed per the product definition. An ideal team for productizing AI research would comprise scientists and engineers with deep AI expertise (in areas such as machine learning and language understanding), software engineers with extensive product knowledge, product managers to define and evangelize the user need and product goals, and user experience designers to bring the work to life inside the product. A critical part of the product manger's role is aligning the group on a shared product vision that spans their organizations. This will ensure the team builds the correct framework and technology to support the longer-term vision while diligently focusing on delivering the minimal viable product to market. In shipping the commitments feature, we assembled a multi-disciplinary team matching the profile described above, spanning product and research. The value proposition for users was clearly understood by team members and was used to ground all conversations and investments. The need for the team to focus on the user and the user experience can be challenging, especially for teams and organizations who are technology first. Having the team aligned on the user instead of only the technology helped ensure that all the investments accrued to product value for customers.

3. Research ready does not mean production ready; learn fast and iterate

The science being "done," and perhaps even published in research literature, doesn't mean that the technology is ready to be integrated into a product and released to users. It is dangerous to estimate time to market without a clear understanding of the release criteria and a detailed assessment of what is required to ship the right user experience for customers.

Existing research is often a demonstration of the feasibility of an idea, but considerable work remains in transferring research findings to practice. This is especially true if there has been a long delay between research and productization. For example, in the commitments project, the gap between the quality and runtime efficiency of the model's output from the prior research versus where they needed to be for a minimum viable product was significant. The fastest way to course correct and determine a realistic timeline for shipping is to build a prototype and get the technology in the hands of real users as quickly as possible, so you can learn. Building tools and prototypes early has many benefits: they can be used to measure quality and quantify the current state, collect data (which is the lifeblood of AI), set exit criteria, measure progress, and communicate your vision and the opportunity to your team, potential investors and other stakeholders. For many it is difficult to understand the full opportunity that AI can offer until they experience it for themselves. In building the commitments feature, we quickly learned that examples in PowerPoint presentations didn't effectively communicate the full opportunity and magic of the experience to our leadership. We decided to build a web-based tool that enabled anyone to see the technology at work on their own email. This was a turning point for the project. The tool enabled our team members, partners, and leaders to experience the technology and value add for themselves, helping them more quickly grasp the opportunity for user delight. Through this we were able to secure investment and collect significant amounts of data to improve our extraction models and work towards setting our release criteria.

When developing AI experiences there is often a need for voluminous and diverse data. You should plan for extended trials with a large representative audience to gather initial user feedback data so that you can test your assumptions about the experience, improve your technology and prepare for release to market. The user experience should be designed with feedback in mind. In shipping the commitments feature, we used an internal beta program to collect qualitative and quantitative data from over two thousand beta users. These data were used to improve the user experience and tune the commitment

extraction models. Thanks to the breadth of feedback data collected through these initial releases, we identified patterns that our commitments extraction model didn't properly handle. For example, we quickly learned that it incorrectly identified statements such as "I will be 10 minutes late to the meeting" as a user commitment, when users do not perceive this as a task that they need to track or remember to do. Refinements were made to exclude those and similar cases. We also performed numerous iterations on the user experience to better present commitments and integrate rich context from the original email thread. As we made these changes and saw increased satisfaction and fewer issues, we decided to extend to the [Windows Insiders Program](#) (WIP), comprised of a larger group of external users who had agreed to receive beta products. The release to this wider audience moved us one step closer to our true user demographic and helped to quantify the gap between where the technology was and where it needed to be to ship to all Cortana users. This helped us determine a more accurate release timeline that we then worked towards, as well as making the value/feasibility of shipping the experience clear to our leadership/investors. The WIP release also provided an early opportunity to test our channels for both collecting feedback and integrating it into our model training pipeline. We increased our agility by releasing to these audiences using a temporary architecture which we knew would not scale to all Cortana users. This helped us learn early from representative users, so we could perfect the AI technology and user experience while the remainder of our engineering team built the architecture needed to ship the experience at scale in parallel. It is critical to plan for this type of extended learning from an early adopter audience early on to set your project up for success.

It is worth noting that we are not alone in adopting an iterative release approach to shipping AI technology. For example, Google took a similar approach with their [Smart Reply](#) feature that helps users quickly compose email replies. After extended time testing the experience internally with a limited audience, they released it externally to users of their Inbox app, where they learned from usage by a smaller early-adopter audience before shipping the technology in their flagship Gmail app.

4. Shipping is only the beginning

Shipping is a huge accomplishment, but it does not signal product completion: continued learning is a vital part of productizing AI. We use implicit and explicit feedback data from users extensively in Cortana today: to personalize experiences, improve our AI and identify product gaps. These data are fed into AI algorithms to help improve their performance over time. The user experience in Cortana commitments was purposely designed to capture user feedback that could be used to improve the feature. Post release is a good time for you to consider extending the signals that are logged beyond the basics that were needed to demonstrate ship readiness. Your product managers need to work closely with your scientists to understand data needs to improve model performance, including what data fields are required, how much data are needed and from whom it should be collected. In fact, since their initial release in Cortana, the commitment extraction models have already gone through several iterations based on user feedback data. Beyond the need to continually refine the AI technology and the experience it powers, other post-shipping challenges you may face include defining an internationalization plan to bring the technology to more markets, integrating it with other applications, and leveraging additional data sources and data types to improve the AI algorithms and their generalizability.

Concluding Remarks

Although there are many challenges in productizing AI research, the differentiation it can bring to products and the utility offered to users (e.g. time and effort saved) can be considerable. Given the wealth of AI research happening in academia and industry, we focused on the process of bringing existing AI research to market. As discussed, it is critical to start by focusing on a specific user problem,

build a team of experts with diverse skill sets, establish the right channels to collect feedback and data early, and ship with the right mechanism in place to continue improving the technology and experience. The lessons we learned and processes we developed from shipping the commitments feature in Cortana have been invaluable to our teams and have paved the way for even more powerful AI experiences to delight our users.

Biographies

Nick Ghotbi is a program manager in Cortana at Microsoft who led the product definition and execution of the commitments feature. His experience is centered around delivering intuitive user experiences augmented by cutting edge machine learning and AI to make users more productive.

Ryen White leads Cortana Research, the applied science team for Cortana at Microsoft, working on next-generation digital assistants. While in Microsoft Research, he led the team that developed the AI technology for the commitments feature.

References

1. Spector, A, Norvig, P, Petrov, S. (2012). Google's hybrid approach to research. *Communications of the ACM*, 55(7), 34-37.