Heilmeier Catechism: Nine Questions To Develop A Meaningful Data Science Project
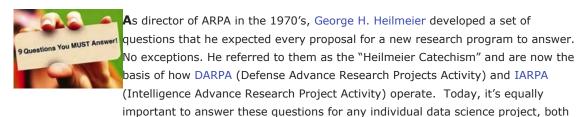
By **Dr. J** - June 11, 2013

As director of ARPA in the 1970's, George H. Heilmeier developed a set of questions that he expected every proposal for a new research program to answer. No exceptions. He referred to them as the "Heilmeier Catechism" and are now the basis of how DARPA (Defense Advance Research Projects Activity) and IARPA (Intelligence Advance Research Project Activity) operate. Today, it's equally important to answer these questions for any individual data science project, both for yourself and for communicating to others what you hope to accomplish.

While there have been many variants on Heilmeier's questions, I still prefer to use the original catechism to guide the development of my data science projects:

1. What are you trying to do? Articulate your objectives using absolutely no jargon. **2**. How is it done today, and what are the limits of current practice? **3**. What's new in your approach and why do you think it will be successful? **4**. Who cares? **5**. If you're successful, what difference will it make? **6**. What are the risks and the payoffs? **7**. How much will it cost? **8**. How long will it take? **9**. What are the midterm and final "exams" to check for success?

Each question is critical in the success chain of events, but number 3 and 5 are most aligned to the way business leaders think. Data science is fought with failures, by the definition of science. As such, business leaders are still a bit (truthfully – a lot) suspicious of how data science teams do what they do and how their results would integrate into the larger enterprise in order to solve real business problems. Part of the data science sales cycle, addressed by question 3, needs to address these concerns. For example, in the post "Objective-Based Data Monetization: A Enterprise Approach to Data Science (EDS)," I present a model for scaling out the our results.

In terms of the differences a project makes (question 5), we need to be sure to cover the business as well as technical differences. The business difference are the standard three: impact on revenue, margin (combined ratios for insurance), and market share. If there is not business value (data/big data economics), then your project is a sunk cost that somebody else will need to make up for.

Here is an example taken from a project proposed in the insurance industry. Brokers are third party entities that sell insurance products on behalf of a company. They are not employees and often are under the governance of underwriters (employee that sells similar products). There are instances where brokers "shop" around looking get coverage for a prospect that might have above average risk (e.g., files too many claims, in high risk business, etc.). They do this by manipulating answers to pre-bind questions (prior to issuing a policy) in order to create a product that will not necessarily need underwriter review and/or approval. This project is designed to help stop this practice, which would help the improve business financial fundamentals. Here is Heilmeier's Catechism for the Pre-Bind Gaming Project:

1. What are you trying to do? Automate the identification of insurance brokers that use corporate policy pricing tools as a means to undersell through third party providers.

2. How is it done today? Corporate underwriters observer broker behaviors and pass judgement based on person criteria.

3. What is new in your approach? Develop signatures algorithms, based on the analysis of gamer/no gamer pre-bind data, that can be implemented across enterprise product applications.

4. Who cares? Business executives – CEO, President, CMO, and CFO.

5. What difference will it make? In an insurance company that generates $350 M in premiums at a combined ratio (margin) of 97%, addressing this problem could result in an additional $12M to $32M of incremental revenue while improving the combined ratio to 95.5%.

6. What are the risks and payoffs? Risks – Not having collect or access to relevant causal data reflecting the gamers patterns. Payoffs – Improved revenue and combined ratios.

7. How much will it cost? Proof of concept (POC) will cost between $80K and $120K. Scaling the POC into the enterprise (implementing algorithms into 5 to 10 product applications) will cost between $500K and $700K.

8. How long will it take? Proof of concept (POC) will take between a 8 to 10 weeks. Scaling the POC into the enterprise will take between 3 to 7 months.

9. What are the midterms & final check points for success? The POC will act as the initial milestone that demonstrates gaming algorithms can be identify with existing data.

Regardless of whether you use Heilmeier's questions or other research topic development methodologies (e.g., The Craft of Research), it is important to systematically address the who, what, when, where, and why of the project. While a firm methodology does not guarantee success, not addressing these nine questions are sure to put you on a risky path, one that will need work to get off of.

Share this:

  Twitter   Facebook

Like this:

Dr. J

Using Data Sciences and Artificial Intelligence to Explore the Darkest Places on Earth - Data