Why are so many AI systems named after Muppets?

One of the biggest trends in AI recently has been the creation of machine learning models that can generate the written word with unprecedented fluidity. These programs are game-changers, potentially supercharging computers' ability to parse and produce language.

But something that's gone largely unnoticed is a secondary trend — a shadow to the first — and that is: a surprising number of these tools are named after Muppets.

To date, this new breed of language AIs includes an ELMo, a BERT, a Grover, a Big BIRD, a Rosita, a RoBERTa, at least *two* ERNIEs (three if you include ERNIE 2.0), and a KERMIT. Big tech players like Google, Facebook, and the Allen Institute for AI are all involved, and the craze has global reach, with Chinese search giant Baidu and Beijing's Tsinghua University contributing models. The naming convention is so well established that these systems are sometimes referred to as "Muppetware." But who started the convention and why?

As you might have guessed, the simple answer is: it's an inside joke, with researchers naming AI models after Muppets because *other* researchers have named AI models after Muppets. But it's a joke that happens to highlight a particular characteristic of AI research, demonstrating how labs pay homage to and build upon one another's work.

> *2018: Language model papers have to introduce Sesame Street-related acronyms*
>
> *2019: Language model papers need Sesame Street jokes in the title, all talks need at least one Sesame Street image.*
>
> *2020: ACL/NAACL co-located with Sesame Street convention, Big*

> *Bird gives a keynote.*
>
> *— Miles Brundage (@Miles_Brundage) June 11, 2019*

The trend started with ELMo, a model devised by the Allen Institute and first published online in October 2017. As is often the case with research that breaks new ground, the team behind the work wanted to come up with a snappy acronym for their model. The paper's lead author, Matt Peters, told *The Verge* over email that they brainstormed ideas on Slack.

"We had a list of letters usable in an acronym," says Peters. "Language Model, Contextual, Embeddings, etc." It was an engineer named Joel Grus who came up with "ELMo" to stand for "Embeddings from Language Models," he says, and the name "instantly stuck."

"My oldest son was about three at the time and it was also my way of dedicating the paper to him."
"I liked it because it is somewhat whimsical but memorable," says Peters. "My oldest son was about three at the time and it was also my way of dedicating the paper to him."

ELMo might have been a one-off had it not been for BERT — a language model created by Google's AI team in 2018. This model proved to be powerful and influential, and pushed a number of novel ideas about language generation into the AI mainstream.

BERT itself officially stands for Bidirectional Encoder Representations from Transformers, and although Google refused multiple requests from *The Verge* to discuss the origins of the name, it's widely assumed that the researchers, like those from Allen, had the Muppets in mind. In Google's own blog post on the topic, the company says "BERT builds upon recent work in pre-training contextual representations — including … ELMo."

BERT achieved state-of-the-art results on a number of tests,

and has been so successful that Google recently incorporated it into its search engine. Once the model was released, the floodgates of Muppetware opened, and it was soon followed by many clever algorithms sporting brute-force acronyms, including ERNIE (Enhanced Representation through Knowledge Integration), KERMIT (Kontextuell Encoder Representations Made by Insertion Transformations), and Big BIRD (Big Bidirectional Insertion Representations for Documents).

Muppetware is having a big impact: Google incorporated BERT into its search engine to improve language comprehension. Credit: Google

But the trend is more than just a joke. As Oren Etzioni, CEO of the Allen Institute, explains, it's also a serious way to recognize "intellectual debt" within the AI world. "ELMo was named thus as a whim, but BERT builds directly on the insights of ELMo; Grover utilizes BERT, etc.," Etzioni told *The Verge* over email. "Emphasizing the credit that is due to ELMo is very important to us … Snuffaluffagus can't be far behind!"

Mitchell Stern, a PhD student at Berkeley who helped create KERMIT and Big BIRD, said the naming convention was mostly fun, but it also had a "branding aspect."

Naming models after Muppets is a way to recognize intellectual debt

"Given how widespread this trend has become, people working in this area will naturally recognize new papers containing a *Sesame Street*-themed name," Stern told *The Verge* by email. And while not every AI language model using these new techniques *is* named after a Muppet (OpenAI's well-traveled GPT-2 systems are just such an exception), it's a pretty sure thing that if you see a Muppetware model you know what approaches it's using.

All this, in turn, helps us to understand how the AI world

depends on openness and collaboration to generate and refine ideas. AI isn’t a discipline where lone scientists toil away in the lab at night, pumping electricity through processors, and cackling “It’s aliiiive” over a glowing command line. (Disclaimer: this certainly does happen, but it’s not always the most productive approach.) Instead, advances tend to be iterative and collaborative, with groups of researchers building upon one another’s work and ideas.

And while it’s possible that the Muppetware joke will wear thin pretty soon, until that happens, it’s a fitting tradition. After all, collaboration and respect are exactly the sort of characteristics that *Sesame Street* characters would be proud of.

https://www.theverge.com/2019/12/11/20993407/ai-language-models-muppets-sesame-street-muppetware-elmo-bert-ernie