

Pre-trained & Foundation Models

rationale, usage, examples, cases

$$f(x_i | x_0 \dots x_{i-1}; \theta_{0, \dots, N})$$

$$N = 175,000,000,000$$

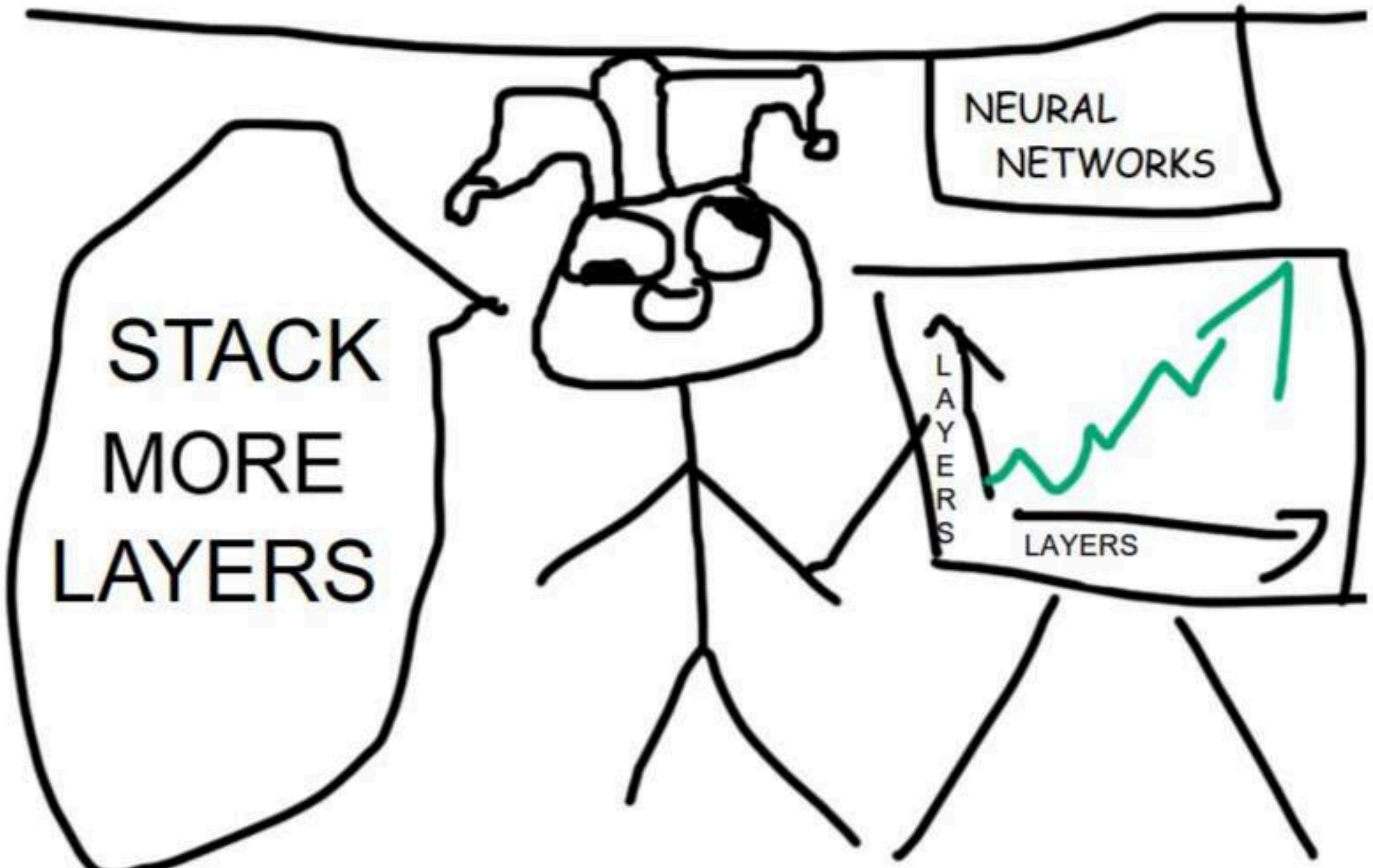
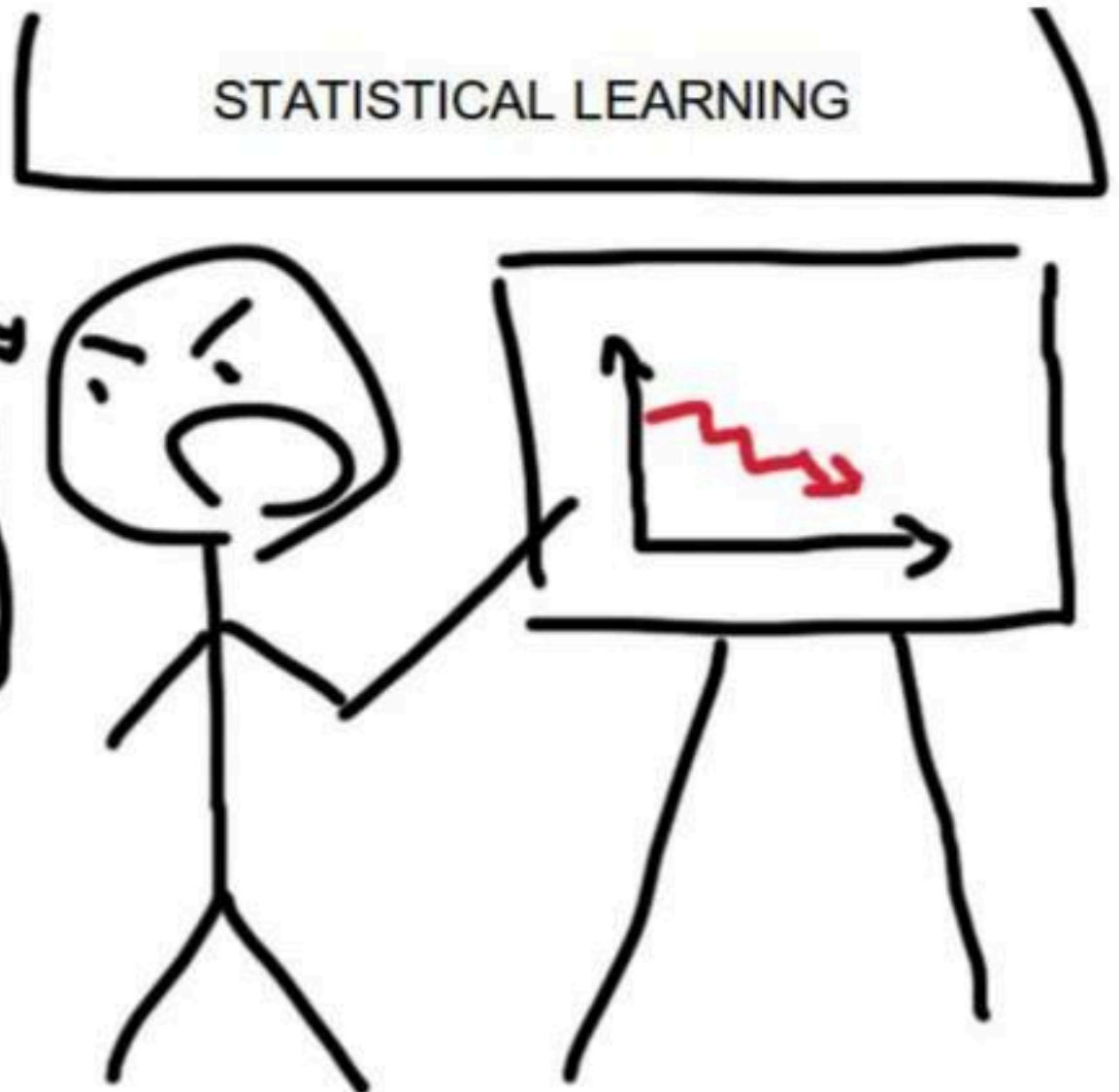
This is for GPT-3.

You learn a function that predicts the next token, and you can sell the parameters of this function.

Good times to be a math geek!

But how?

ladies and
Gentlemen, our learner
overgeneralizes because the
VC-Dimension of our Kernel
is too high, Get some
experts and minimize the
structural risk in a new one.
Rework our loss function,
make the next kernel stable,
unbiased and consider using a
soft margin



How?

- Complex architectures (e.g., the transformer architecture)
- Massive data
- Learning algorithms to handle massive amount of data

All contributed.

Pre-trained models—initial examples

- Pretrained models are intended to generalize well to custom datasets with minimal effort, for a specific task
 - These are pretrained models that we believe would generalize to future data
- They can be partly trained later for finetuning on a new dataset
- Typical example is transfer learning
 - An example usage is to obtain a large image classifier model trained on ImageNet data
 - Remove the output layer
 - Perhaps add a few more layers
 - Add the new output layer for the new objective
 - Freeze original model weights and only train newly introduced layers with new data
- Use the classifier with the hope that the original model we transferred from model will effectively serve as a generic model of the visual world

From pre-trained to "foundation" models

Foundation models are transforming AI.

They are **general**:

They are "foundation models," because:

"AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. **We call these models foundation models to underscore their critically central yet incomplete character.**" *[Stanford paper: "On the opportunities and risks of foundation models"]*

From pre-trained to "foundation" models

They have **few-shot** (even one-shot or zero-shot) capabilities:

- That is, they can perform a wide variety of tasks that they are not specifically trained for (for instance, a language model, which is trained to predict the next word given a context, can perform text categorization, too.) [*OpenAI paper: "Language models are few-shot learners"*]
- They can act as a platform for creating specialized models, too, through fine-tuning—they serve as a base for new models

From pre-trained to "foundation" models

Sometimes they are "**conversational**," and "**generative**"

- We interact with the model through natural language (we can invent tasks through prompts!)
- They can generate text, images, audio or visual content based on prompts, etc.
- Some are open-source, some are private and hosted inside a company, while others are hosted on cloud platforms and made accessible via APIs

What's in it for you?

Old times

- You need a team of experts
- You need training data
- You need to experiment with models
- Train-test-hope to generalize

This is a tedious process! And it takes a lot of time, effort, and resources.

Using pre-trained models in your projects

- They save us from collecting large datasets
- They save us from training complex architectures from scratch
- Yet, these models are adaptable to similar tasks that we intend we solve
- They give impressive results with minimum effort and compute resources

Using pre-trained models in your projects

But these models need to be

- gated,
- customized,
- engineered,
- composed together,
- integrated,
- tuned,
- augmented

to **solve real business problems**

Our approach at the Ally AI

"AI for fun" solutions are on the run.

However, numerous real business problems are waiting to be addressed.

AI's primary purpose should be to assist businesses in generating tangible value.

We know. We are here to help you do so.

Benefits

- HUGE ROI!
- Rapid results
- Low price point—No more ignoring problems (because of the cost of collecting training data and an gathering an ML team together) that can now be solved due to the generality of foundation models
- Simple—Integrated with tools you use everyday (e.g., ERP, chat, intranet, CRM, email, etc.)

Example use cases/solutions

Active Listening

Actively listen to overall performance of call center operations, from agent performance to customer sentiment.

Call Routing

Direct incoming calls to the most appropriate individual based on the caller's history, query, sentiment, value etc.

Virtual Desktop Agent

Transcribe speech in real-time, comprehend and respond to customer inquiries using natural language processing, analyze sentiment to adjust interactions, initiate follow-up actions, and monitor call quality and agent performance.

Example use cases/solutions

Voice of the Customer

Analyze real-time customer feedback to gauge overall sentiment and customers' potential actions towards products or services.

Voice of the Store

Analyze feedback from employees in-store about products, issues, customer feedback etc. - from any channel and in any data form.

Example use cases/solutions

Catalog Builder

Automate the creation, organization, and unification of online product catalogs, generate relevant descriptions and tags for products.

User Generated Content Moderation

Automatically moderate vast amounts of user generated content, minimizing misleading, fraudulent content on platforms and marketplaces.

Pre-trained & Foundation Models

gokhan@theally.ai