Engineering AI vs. Science AI

Alexei (Alyosha) Efros
UC Berkeley

© Quint Buchholz
Things are changing fast

- When **GPT4 Image** model is released, it will likely render **80%** of this CVPR’s papers obsolete!
  - I’ve been urging to focus on data for 25 years, and even I am surprised!
Not our first rodeo

• 2000s: The great *Geometry Extinction Event*
  – In ICCV 1999, **zero** orals in recognition!
  – By CVPR 2009, it was mostly recognition
  – **Pop quiz:** *in VGGNet, what does “VGG” stand for?*

• 2000s: Rise of Datasets, Mean AP, SOTA
  – Datasets became huge, multi-year, multi-person, high-investment efforts
  – Had to learn to scale up our algorithms
Not our first rodeo

• **2010s:** The Deep Learning Revolution
  – 80% of papers before 2013 became obsolete
  – GPUs are expensive
  – Scaling up is challenging
  – Adapt or Perish
    • Most folks adapted

• **2023:** Billion Image Models
  – Surely we can handle $10^2$ increase in 20 years!
Longer-Term Future of Computer Science (including Computer Vision and AI)

Is CS Science or Engineering?
The Great Decoupling

Engineering AI
- Cool uses of machine learning
  - robo-receptionists, robo-lawyers, self-driving cars, self-building houses, etc.
- Will be mostly done in Industry
  [https://youtu.be/kK4biRfcwxY](https://youtu.be/kK4biRfcwxY)

Science AI
- Drift closer to natural sciences
  - Evolutionary biology, cognitive science, developmental psychology, anthropology, philosophy
- Too long-horizon for Industry
  Alexei Efros: Learning From the Ground Up, The Batch, 2021
Science AI to Understand Intelligence to use computational tools to understand and perhaps model complex processes governing biological organisms, their societies, whole ecosystems, maybe even evolution itself.
Two Ingredients for True AI:

1. Focus on Data (over Algorithms)
2. Emergent Objectives
Algorithms vs. Data

For long time, data didn’t get much love…

Data
Features
Learning Algorithm
Algorithms vs. Data (late 1990s)

- Rowley, Baluja, and Kanade, 1998
  - features: pixels, classifier: neural network
- Schniederan & Kanade, 1999
  - features: pairs of wavelet coeff., classifier: naïve Bayes
- Viola & Jones, 2001
  - features: haar, classifier: boosted cascade
Algorithms vs. Data (2023)

<table>
<thead>
<tr>
<th>Diffusion-based</th>
<th>Auto-regressive</th>
<th>GAN-based</th>
</tr>
</thead>
</table>

Prompt: “squirrel reaching for a nut”
Algorithms vs. Data (2023)
Our Scientific Narcissism

All things being equal, we prefer to credit our own cleverness
Two Ingredients for True AI:

1. Focus on **Data** (over Algorithms)

2. **Emergent Objectives**
   - Data is crucial – it grounds us in the world, but alone it is just for mimicry and pastiche
     - Sure, GPT-X will answer 95% of your e-mails
     - Will write bad poetry, mass-market romance novels.
Emergence vs. Copying


15 Years Later
True Intelligence Must Emerge

“AI is not when computer can write poetry, AI is when computer will \textit{want} to write poetry”

-- young soviet physicist in the 1980s
Will Engineering AI create Borges or Bach?

- Creativity is not an objective
- it’s a byproduct of all the intricacies and complexities of the human condition
- Grounded in the world for a billion years. No shortcuts!
Long Road Ahead

Understanding and modeling the processes governing the slow and steady emergence of what we call intelligence, from bacteria to Bach, is the great long-term challenge for Science AI.