50.039 Theory and Practice of Deep Learning

W12S3 – End and Review

Matthieu De Mari

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Some admin stuff

Final exam

- Content will be everything from W1 to W12 included.
- Similar format as MidTerm.
- Details for exam (location, assignments, etc.) to be sent via email.

Regarding your project

- **No extension will be given (have to give grades to OSA!)**
- **And want to be able to see your projects before presentations and discussions on W13.**
- **Check schedule for presentation timeslots.**

So this is the end

What’s next?

Develop a deeper understanding of…

Improving training procedures (W2++)

At the moment, mostly using gradient descent algorithms to train our models…

Many different directions have been considered, for instance:

- Using Forward-Forward (2022 proposal from Hinton, to replace our conventional backprop?): https://arxiv.org/abs/2212.13345
- Training AIs to train other AIs?: “learning to learn” or “meta-learning”
  Curious?: https://machinelearningmastery.com/meta-learning-in-machine-learning/

On top of everything we have seen…

Advanced Computer Vision (W4++ and W10++)

Consider enrolling for the Computer Vision Term 7 course for more advanced concepts on CV, such as:

- More advanced loss functions like triplet loss,
- Advanced architectures like Siamese networks,
- Video data models,
- Etc.

https://istd.sutd.edu.sg/undergraduate/courses/50035-computer-vision

Develop a deeper understanding of…

Advanced attacks and Defense mechanisms (W5++)

Many more mechanisms when it comes to attacking and defending a Neural Network, e.g. new types of attacks, such as:

- Poisoning attacks (attempt to poison the dataset so the NN cannot re-train properly, especially if online learning is used),
- Weights changes (attempt to change a small subset of the weights of the NN to prevent it from working in certain ways),
- Etc.

https://www.comp.nus.edu.sg/~reza/courses/cs6231/

Develop a deeper understanding of…

Advanced word embedding and NLP problems (W6++ and W8++)

Many more mechanisms when it comes to embedding and language related problems.

- E.g. more advanced embeddings
- Typical tasks in NLP (chatbots, context propagation, sentiment analysis, translation, etc.)
- Go for the Term 7 NLP course!

https://istd.sutd.edu.sg/undergraduate/courses/50040-natural-language-processing

Develop a deeper understanding of…

Advanced Graph Neural Networks (W9++)

We barely scratched the surface of Graph Theory. If you need to study a new math theory, let it be graph theory!

- Good graph theory course here:

https://ocw.mit.edu/courses/mathematics/18-217-graph-theory-and-additive-combinatorics-fall-2019/

- More advanced problems and concepts on Graph Neural Networks in lectures 1-9 of course here:

https://www.cs.ox.ac.uk/teaching/courses/2020-2021/advml/

Develop a deeper understanding of…

Advanced Graph Neural Networks (W9++)

Also, keep in mind that Neural Networks are graphs…

- So technically, we could build a Neural Network, which receives another Neural Network as its input…!
- What could be the uses for such a technique?
- Meta-learning? (i.e. training an AI to train another AI?!) https://machinelearningmastery.com/meta-learning-in-machine-learning/

Develop a deeper understanding of…

Advanced Generative Models (W10++)

- Advanced GANs, operating on other types of data than just images (sound, text, etc.)
- Very good online course here:

https://cs236g.stanford.edu/

- Any good course about advanced diffusion models and advanced generative models (Dall-E and MidJourney) would also be worth considering.

https://www.fast.ai/posts/part2-2023.html

- Might be covered in **Term 7 Computer Vision**?

Develop a deeper understanding of…

Advanced Reinforcement Learning (W11++)

Barely scratched the surface about Reinforcement Learning.

- Currently considering to create a Reinforcement Learning elective course at SUTD for Term 8. Any thoughts?
- Otherwise, the reference course on RL is the one from David **Sliver** (the man behind AlphaGo!)

https://deepmind.com/learning-resources/-introduction-reinforcement-learning-david-silver

And https://www.davidsilver.uk/teaching/

Develop a deeper understanding of…

Advanced Interpretability (W12++)

- Rather an ongoing field in research at the moment.
- Not that many courses out there, but worth keeping an eye out…

On top of everything we have seen…

More concepts, problems and architectures on Computer Vision

- Also, always good to go for an image processing course to understand typical image transformation and problems out there.

https://www.coursera.org/learn/image-processing?ranMID=40328&ranEAID=*GqSdLGGurk&ranSiteID=.GqSdLGGurk-GV4LxEnPMuMd1.8y4AurRA&siteID=.GqSdLGGurk-GV4LxEnPMuMd1.8y4AurRA&utm_content=10&utm_medium=partners&utm_source=linkshare&utm_campaign=*GqSdLGGurk

On top of everything we have seen…

Bayesian and Statistical Learning (Variational AutoEncoders were 101, more on diffusion models).

- A good entry point for Bayesian Deep Learning

https://medium.com/@ODSC/introduction-to-bayesian-deep-learning-f7568f524c90

- Lectures 10-End

https://www.cs.ox.ac.uk/teaching/courses/2020-2021/advml/

On top of everything we have seen…

A bit of advanced optimization and game theory never hurts…

- Especially when trying to optimize two cooperating or competing neural networks! (GANs, actor-critic, etc.)
- Great courses here:

https://oyc.yale.edu/economics/econ-159

And

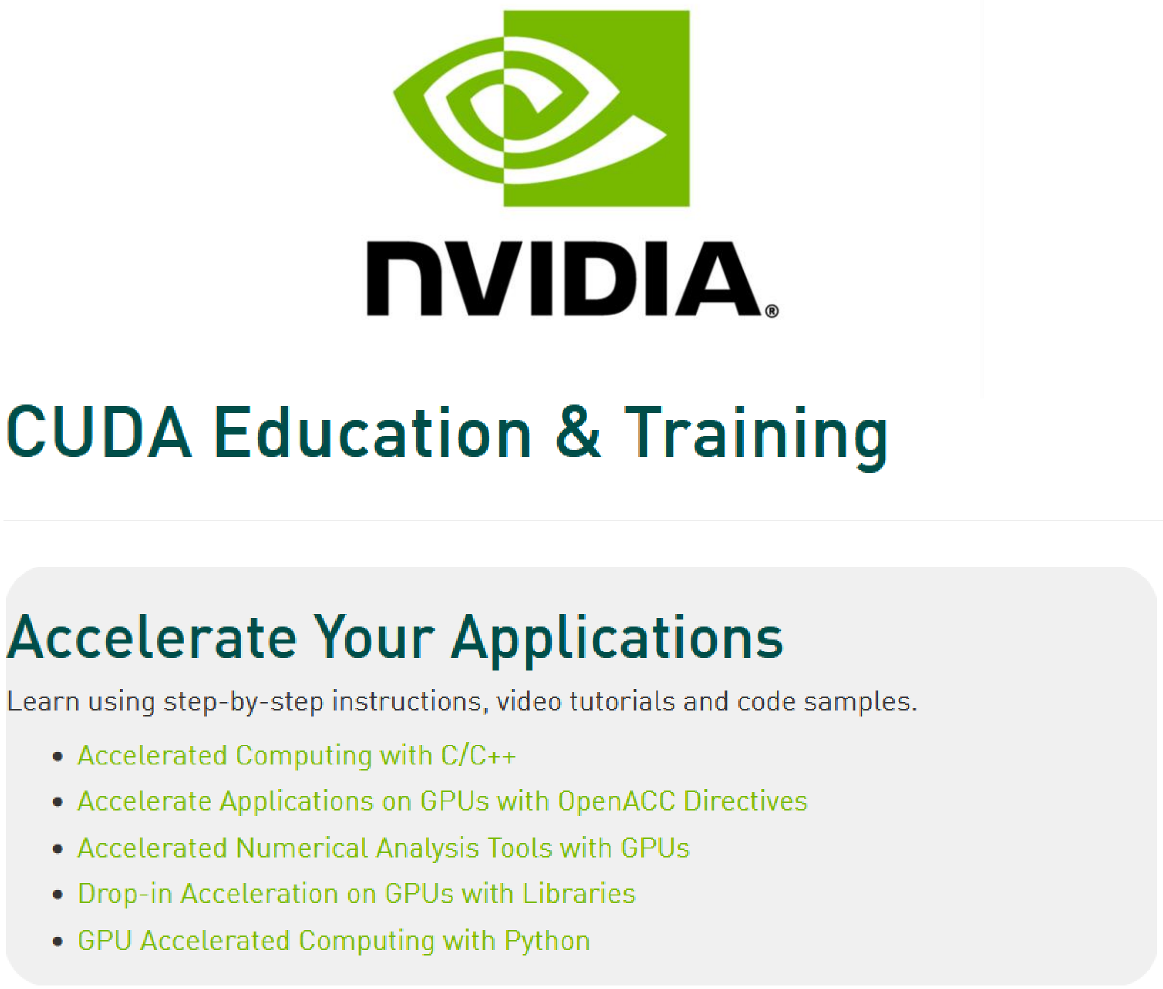
https://online.stanford.edu/courses/soe-ycs0002-game-theory

On top of everything we have seen…

CUDA masters are the king of the world these days…

- BigTech companies are looking for experts that can help with machine learning and custom GPU implementations
- The most obvious way to learn is from Nvidia courses themselves, some give certifications, but it is an investment…

https://developer.nvidia.com/cuda-education-training



CUDA Education & Training

Accelerate Your Applications

Learn using step-by-step instructions, video tutorials and code samples.

- Accelerated Computing with C/C++
- Accelerate Applications on GPUs with OpenACC Directives
- Accelerated Numerical Analysis Tools with GPUs
- Drop-in Acceleration on GPUs with Libraries
- GPU Accelerated Computing with Python

On top of everything we have seen…

Could computing is also very valuable…

- Similarly, a certification in AWS or Microsoft Azure of Google Cloud for cloud computing machine/deep learning is of high value these days!

On top of everything we have seen…

Quantum is the next best thing?

Quantum computers are expected to be the next big thing in Computer Science in general.

- This will also apply to AI/ML/DL…
- This means we will get to train larger networks, faster. (This is currently a limit for many applications these days).
- Picking up on quantum computing is never a bad idea (but careful, possibly the most difficult topic out there!)

https://towardsdatascience.com/dont-ask-what-quantum-computing-can-do-for-machine-learning-cc44feeb51e8

https://pennylane.ai/qml/whatisqml.html

On top of everything we have seen…

More stuff

- Advanced Probability and Statistics (a.k.a. Statistical Learning) is always a great plus… https://www.statlearning.com/
- Neuroscience should probably be part of any serious AI curriculum… [NeuroAI] Barron et al., “What insects can tell us about the origins of consciousness”, 2015.
  [Kurzweil] R. Kurzweil, “How to Create a Mind: The Secret of Human Thought Revealed”, 2012.
  [Marcus] G. Marcus, “The Future of the Brain”, 2014.

Etc.

On top of everything we have seen…

More stuff

Many more courses available for free/cheap out there…

- https://www.linkedin.com/posts/endritrestelica_ai-artificialintelliegence-activity-7069768595978776576-Cvsk/
- https://www.linkedin.com/posts/amanc_artificialintelligence-machinelearning-ai-activity-7052863983908753408-OPfM/

Possibly the best repo ever created on GitHub for CS stuff…

- https://github.com/Developer-Y/cs-video-courses

On top of everything we have seen…

More stuff

- Using DL to solve complex differential equations. https://medium.com/swlh/artificial-intelligence-can-now-solve-a-mathematical-problem-that-can-make-researchers-life-easier-9602c869128
- General AI, i.e. designing an AI will full human cognitive capabilities (vision, hearing, speech, movement, etc.). https://www.forbes.com/sites/forbestechcouncil/2021/07/16/the-future-of-artificial-general-intelligence/?sh=c9223323ba99
- Etc.

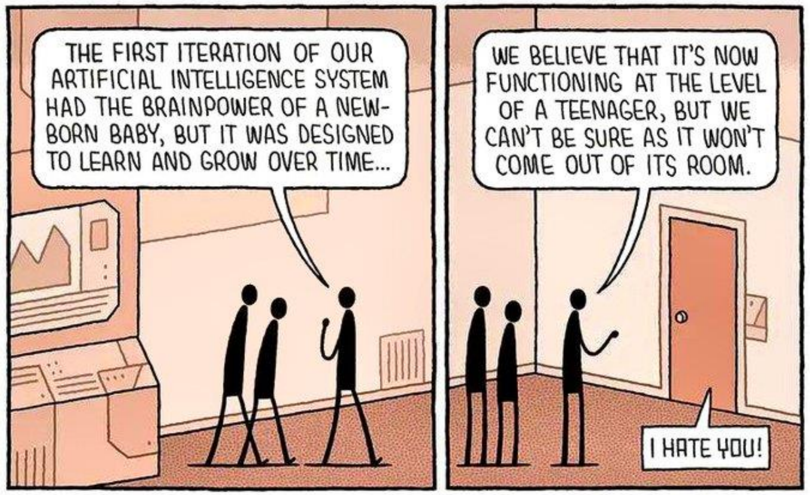
Debates about AGI

AGI: Artificial General Intelligence, naming the idea of an AI that would match all the cognitive abilities of a human.

- At the moment, definitely a dream, but the ultimate goal.
- Very active discussion topic after ChatGPT has been released.
- ChatGPT is nowhere near AGI, but is it a big step in that direction?

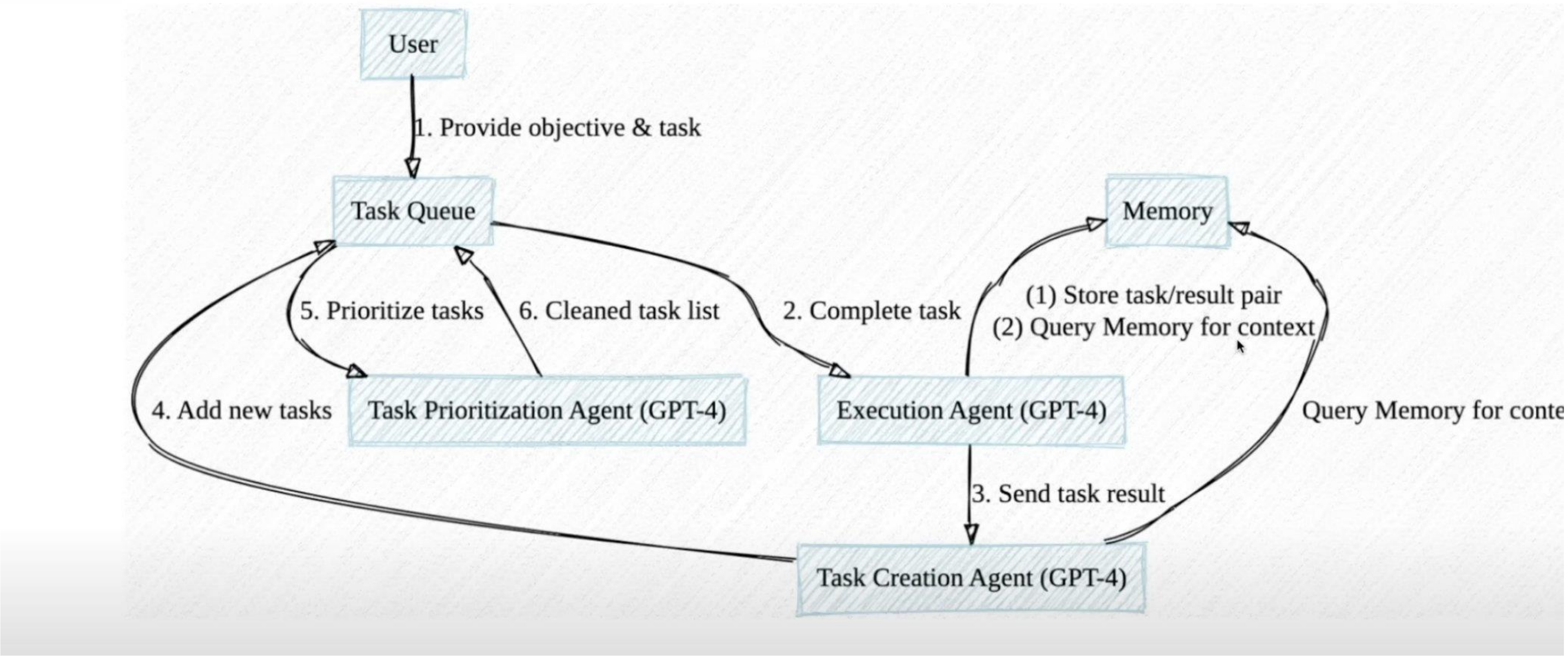
I'm staying out of this debate! (Still watching it unfold though *popcorn*)

https://openai.com/blog/planning-for-agi-and-beyond


THE FIRST ITERATION OF OUR ARTIFICIAL INTELLIGENCE SYSTEM HAD THE BRAINPOWER OF A NEW-BORN BABY, BUT IT WAS DESIGNED TO LEARN AND GROW OVER TIME...
WE BELIEVE THAT IT'S NOW FUNCTIONING AT THE LEVEL OF A TEENAGER, BUT WE CAN'T BE SURE AS IT WON'T COME OUT OF ITS ROOM.
I HATE YOU!

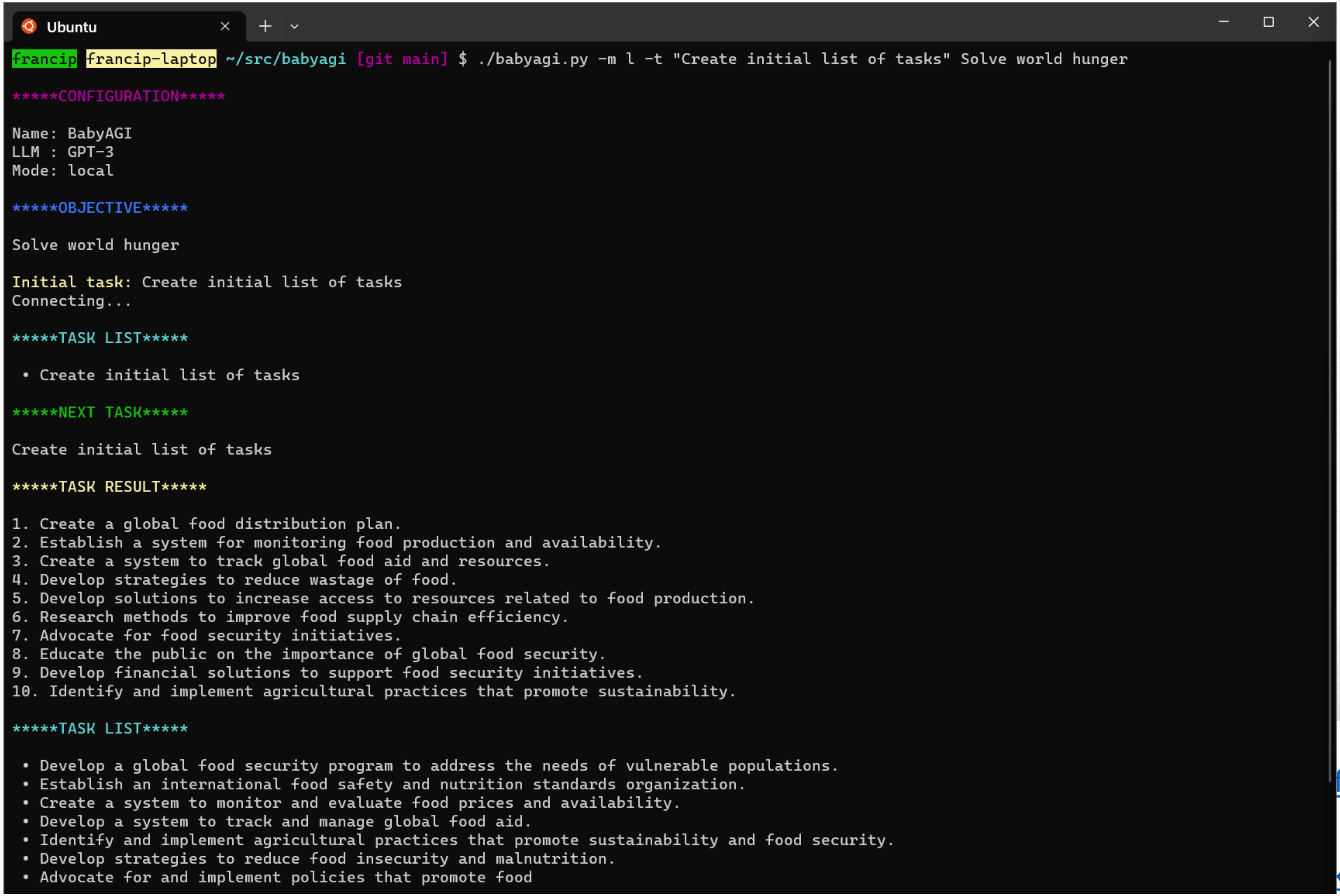
TOM GAULD for NEW SCIENTIST

A good pointer: the babyAGI project

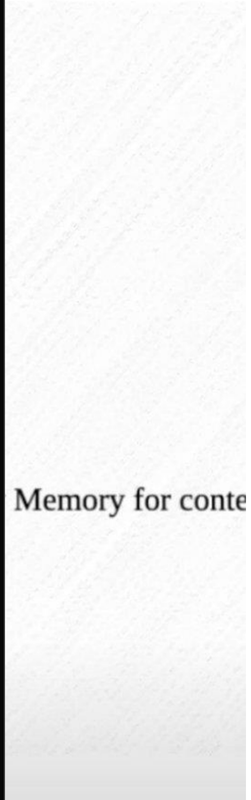


https://twitter.com/yoheinakajima?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor

https://github.com/yoheinakajima/babyagi



```
francip francip-laptop ~/src/babyagi [git main] $ ./babyagi.py -m l -t "Create initial list of tasks" Solve world hunger

*****CONFIGURATION*****

Name: BabyAGI
LLM : GPT-3
Mode: local

*****OBJECTIVE*****

Solve world hunger

Initial task: Create initial list of tasks
Connecting...

*****TASK LIST*****

 • Create initial list of tasks

*****NEXT TASK*****

Create initial list of tasks

*****TASK RESULT*****

1. Create a global food distribution plan.
2. Establish a system for monitoring food production and availability.
3. Create a system to track global food aid and resources.
4. Develop strategies to reduce wastage of food.
5. Develop solutions to increase access to resources related to food production.
6. Research methods to improve food supply chain efficiency.
7. Advocate for food security initiatives.
8. Educate the public on the importance of global food security.
9. Develop financial solutions to support food security initiatives.
10. Identify and implement agricultural practices that promote sustainability.

*****TASK LIST*****

 • Develop a global food security program to address the needs of vulnerable populations.
 • Establish an international food safety and nutrition standards organization.
 • Create a system to monitor and evaluate food prices and availability.
 • Develop a system to track and manage global food aid.
 • Identify and implement agricultural practices that promote sustainability and food security.
 • Develop strategies to reduce food insecurity and malnutrition.
 • Advocate for and implement policies that promote food
```



f_src=twsrc%5Egoogle%7

kajima/babyagi

The important message is…

Your learning should not stop after SUTD…

Keep learning to stay up to date, this is a very fast evolving field…

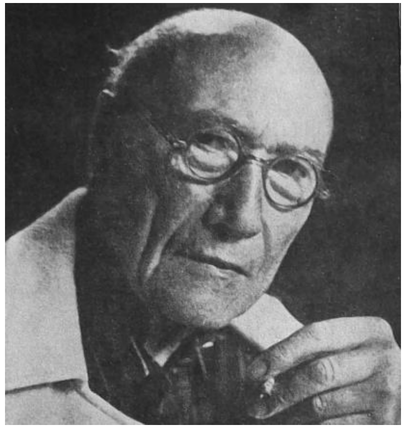
So, good luck on your continuing studies!

More importantly

DL/AI is a very active and fast-paced field.

- Keep your watchlist of papers and authors up to date.
- I have mentioned researchers, which I believe are among the most notable influencers of the Deep Learning community.
- Will be adding some more names on the next slides.

“A good professor should have this constant concern: teaching his students how to continue without him.“
– André Gide, Nobel Prize of Literature in 1947



Add these researchers, companies and researche groupes to your watchlist

- **Demis Hassabis**: Co-founder of **DeepMind**, **AlphaGo**. Several contributions in **Reinforcement Learning**. https://scholar.google.com/citations?hl=en&user=dYpPMQEAAAAJ
- **Alex Graves**: **Professor** at **University of Toronto**. Several contributions in **Reinforcement Learning**. https://scholar.google.co.uk/citations?user=DaFHynwAAAAJ&hl=en
- **Michael I. Jordan**: Professor at **UC Berkeley**, co-inventor of LDA. https://scholar.google.com/citations?user=yxUduqMAAAAJ&hl=fr
- **Terrence Sejnowski**: **Professor** at **UC San Diego**, **Boltzmann machines**. https://scholar.google.ca/citations?user=m1qAiOUAAAAJ&hl=en

Add these researchers, companies and research groupes to your watchlist

- **Peter Norvig**: **Director of Research** at **Google**, co-author of the other Bible of Deep Learning
  https://scholar.google.com/citations?user=Ol0vcWgAAAAJ&hl=en
  http://aima.cs.berkeley.edu/
- **Stuart Russell**: **Professor** at **UC Berkely**, co-author of the other Bible of Deep Learning
  https://scholar.google.com/citations?user=2oy3OXYAAAAJ&hl=en
- **Francois Chollet**: **Researcher** at **Google**. The man behind the **Keras** framework and **Xception**.
  https://scholar.google.com/citations?user=VfYhf2wAAAAJ&hl=en

Add these researchers, companies and research groupes to your watchlist

- **Trevor Hastie**: **Professor** at **Stanford**, co-autor of the Bible of Statistical Learning.
  https://scholar.google.ca/citations?user=tQVe-fAAAAAJ&hl=en
  https://hastie.su.domains/ElemStatLearn/download.html
- **Robert Tibshirani**: **Professor** at **Stanford**, co-autor of the Bible of Statistical Learning. Inventor of the LASSO algorithm.
  https://scholar.google.ca/citations?user=ZpG_cJwAAAAJ&hl=en
- **Vladimir Vapnik: Retired Professor**, inventor of **SVMs** and many other concepts. Worked with **Yann LeCun** at **Facebook AI**.
  https://scholar.google.com/citations?user=vtegaJgAAAAJ&hl=fr

Add these researchers, companies and research groupes to your watchlist

- **Fred Cummins**: **Professor** at **University College Dublin**, contributions to **LSTMs** and **NLP**.
  https://scholar.google.com/citations?user=E-vg2zQAAAAJ&hl=fr
- **Andrej Karpathy**: **Former Director of AI** at **Tesla**. Many contributions to Computer Vision (**Imagenet**) and NLP (**RNNs**).
  (Probably better to follow him than Elon Musk.)
  https://scholar.google.com/citations?user=l8WuQJgAAAAJ&hl=fr
- **Li Fei-Fei**: **Professor** at **Stanford**. Many contributions to Computer Vision (**Imagenet**).
  https://scholar.google.com/citations?user=rDfyQnIAAAAJ&hl=fr
- **Pieter Abbeel**: **Professor** at **UC Berkeley**, and a leading researcher in reinforcement learning and robotics.
  https://scholar.google.com/citations?user=vtwH6GkAAAAJ&hl=en

Add these researchers, companies and research groupes to your watchlist

- **Anil K. Jain**: **Professor** at **Michigan State University**. Many contributions to Computer Vision and Statistical Learning. https://scholar.google.com/citations?user=g-_ZXGsAAAAJ&hl=fr
- **Jitendra Malik**: **Professor** at **UC Berkeley**. Many contributions to Computer Vision and Statistical Learning. https://scholar.google.com/citations?user=oY9R5YQAAAAJ&hl=fr
- **Sebastian Thrun**: **Stanford**, cool stuff on **robotics**. https://scholar.google.com/citations?user=7K34d7cAAAAJ&hl=fr
- **Daphne Koller**: **CEO** at **InSitro**, some cool **courses on Coursera**, she might be the co-founder of Coursera (?). https://scholar.google.com/citations?user=5Iqe53IAAAAJ&hl=en

Add these researchers, companies and researchs groupes to your watchlist

- **Andrew Ng**: **Professor** at **Stanford**, co-creator of **Coursera**. Has one of the best online courses on Deep Learning. https://scholar.google.com/citations?user=mG4imMEAAAAJ&hl=en
- **Jeremy Howard**: **Research Scientist** at **University of San Francisco**, a good scout for notable research papers on Twitter and **TED talks**. https://scholar.google.com/citations?user=ZWdEJ54AAAAJ&hl=en
- **Yaser S. Abu-Mostafa**: **Professor** at **CalTech**, one of the best professors for Deep Learning out there. https://dblp.org/pid/69/3008.html
- **Rachel L. Thomas**: **University** of **San Francisco**, **FastAI**, some great TED conferences on AI and Deep Learning. https://scholar.google.com/citations?user=BDsAYUsAAAAJ&hl=en

Also worth subscribing to a few free newsletters about AI/DL

- The Batch newsletter by DeepLearning.ai (https://www.deeplearning.ai/the-batch/)
- The Algorithm by MIT Tech Review (https://www.technologyreview.com/newsletter-preferences/)
- The TLDR; newsletter (https://tldr.tech/)
- The NLP Newsletter (https://www.ruder.io/nlp-news/)
- The AlphaSignal newsletter (https://alphasignal.ai/)

Some Medium subscription never hurts (sometimes nice, easy and accessible discussions about AI).

But most seriously though, go for Twitter and follow people.

Cracking the Deep Learning interview

Most BigTech companies out there will test you on your CS/DL skills upon applying to positions/jobs in their companies.

- Train!
- Many flashcards with typical AI/ML/DL questions and tasks given in technical interviews at BigTech companies.
- https://www.algoexpert.io/machine-learning/product
- https://www.mlexpert.io/

Cracking the Deep Learning interview

More importantly, polish your GitHub with some nice ML/AI/DL projects that you can show and discuss in interviews!

- Document your codes and notebooks over the summer break!
- Many problems out there, easily investigated.
- https://twitter.com/0xbnomial/status/1523256409529667584?t=YWXdKryCUrrJSCKWLiT4Tg&s=03

- Other possible idea: Find one problem for each concept we discussed each week this term and make a repo/project about it!

A quick word on PyTorch 2.0

PyTorch 2.0, released on 23 March 2023.

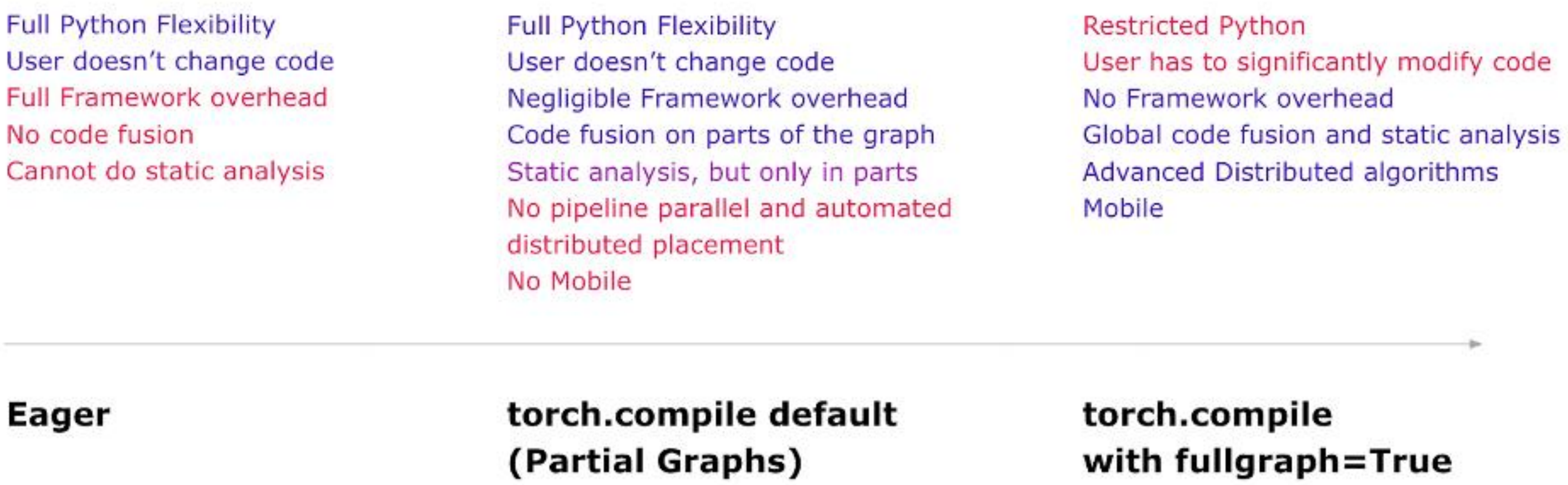
- If you go for it, make sure to correctly set your CUDA (will most likely not be the same version of CUDA if you were using 1.13 before!)
- https://pytorch.org/get-started/pytorch-2.0/

New and remarkable features:

- Fully backward compatible (all previous codes should still work).
- More advanced compile function (three modes) to transform model written in Python into something very fast to execute.
- (More primitive functions and user experience tools.)

A quick word on PyTorch 2.0

More advanced *compile()* function (three modes) to transform model written in Python into something very fast to execute.



Full Python Flexibility
User doesn’t change code
Full Framework overhead
No code fusion
Cannot do static analysis

Eager

Full Python Flexibility
User doesn’t change code
Negligible Framework overhead
Code fusion on parts of the graph
Static analysis, but only in parts
No pipeline parallel and automated distributed placement
No Mobile

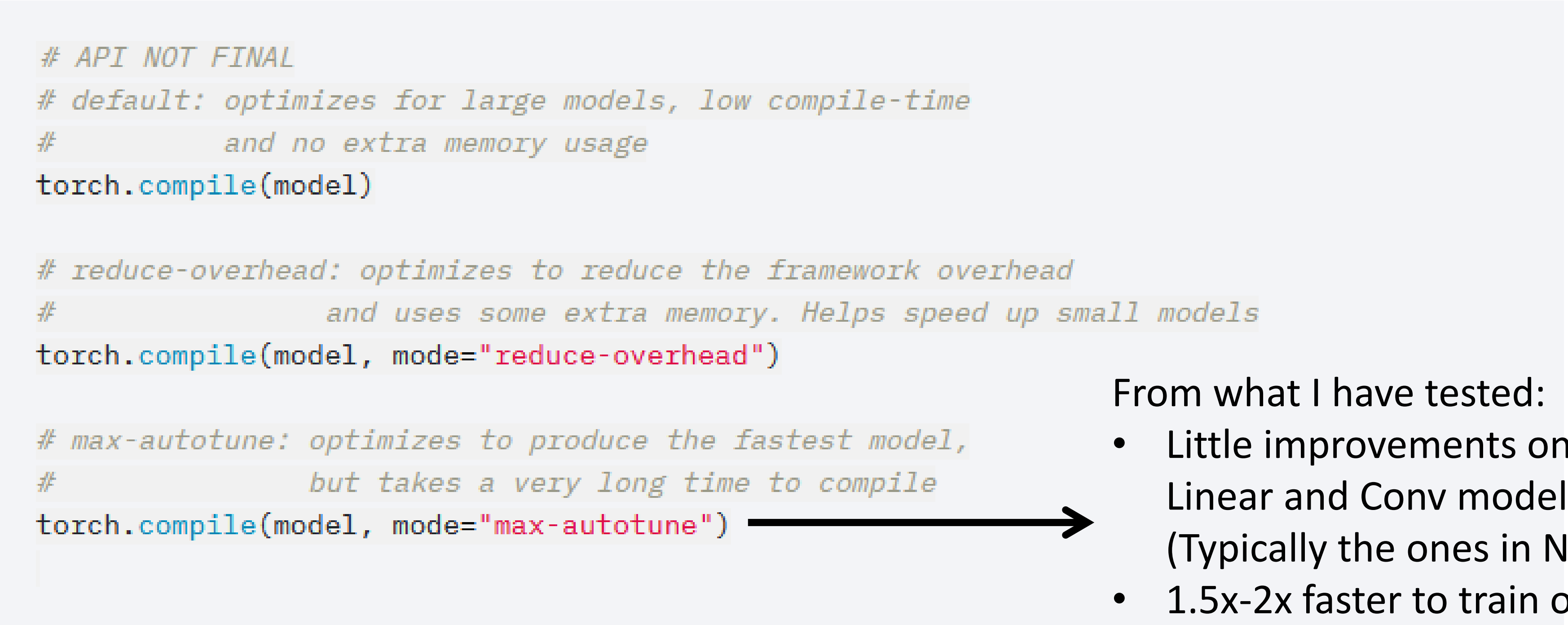
torch.compile default (Partial Graphs)

Restricted Python
User has to significantly modify code
No Framework overhead
Global code fusion and static analysis
Advanced Distributed algorithms
Mobile

torch.compile with fullgraph=True

A quick word on PyTorch 2.0

More advanced *compile()* function (three modes) to transform model written in Python into something very fast to execute.



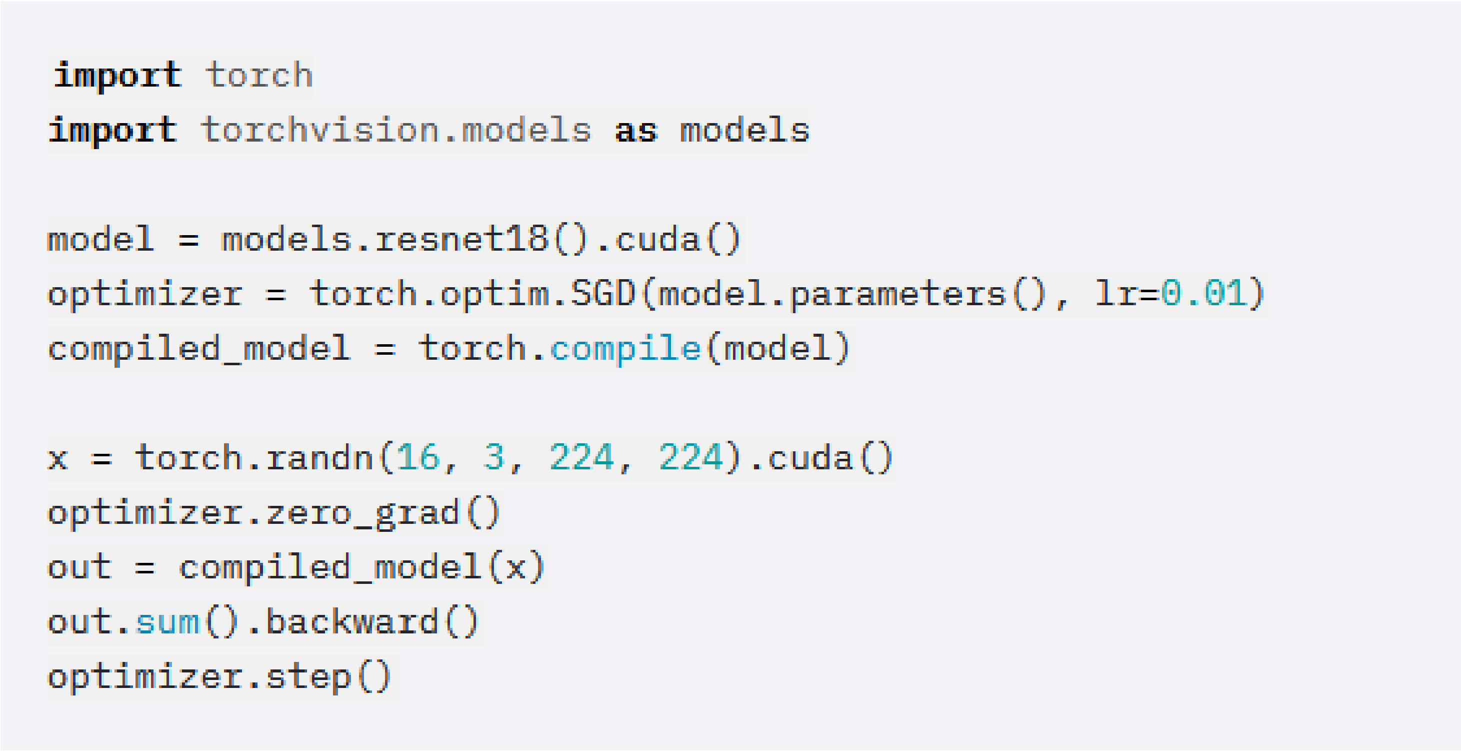
```
# API NOT FINAL
# default: optimizes for large models, low compile-time
#          and no extra memory usage
torch.compile(model)

# reduce-overhead: optimizes to reduce the framework overhead
#                and uses some extra memory. Helps speed up small models
torch.compile(model, mode="reduce-overhead")

# max-autotune: optimizes to produce the fastest model,
#               but takes a very long time to compile
torch.compile(model, mode="max-autotune")
```

From what I have tested:

- Little improvements on small sized Linear and Conv models. (Typically the ones in Notebooks)
- 1.5x-2x faster to train on large-scale transformer architectures?

A quick word on PyTorch 2.0

More advanced *compile()* function (three modes) to transform model written in Python into something very fast to execute.



```
import torch
import torchvision.models as models

model = models.resnet18().cuda()
optimizer = torch.optim.SGD(model.parameters(), lr=0.01)
compiled_model = torch.compile(model)

x = torch.randn(16, 3, 224, 224).cuda()
optimizer.zero_grad()
out = compiled_model(x)
out.sum().backward()
optimizer.step()
```

From what I have tested:

- Little improvements on small sized Linear and Conv models. (Typically the ones in Notebooks)
- 1.5x-2x faster to train on large-scale transformer architectures?