Decentralized Orchestration

for training Deep Neural Networks

Abhishek Singh

PhD Student
MIT Media Lab

Berkeley
UNIVERSITY OF CALIFORNIA

Intelligence = Scale × Heterogeneity

Brown et al., Language Models are Few-Shot Learners, NeurIPS'20

Berkeley
UNIVERSITY OF CALIFORNIA

Heterogeneity

With Heterogeneity

- Mitigate Bias
- Improve Robustness
- Model heterogeneity with MoE, Multimodal learning etc.

Without Heterogeneity

- Mode Collapse
- Neural Collapse
- Poor generalization

1. Bukharin, A., & Zhao, T. (2023). Data diversity matters for robust instruction tuning. arXiv preprint arXiv:2311.14736.
2. Brown et al., Language Models are Few-Shot Learners, NeurIPS'20

Heterogeneity

With Heterogeneity

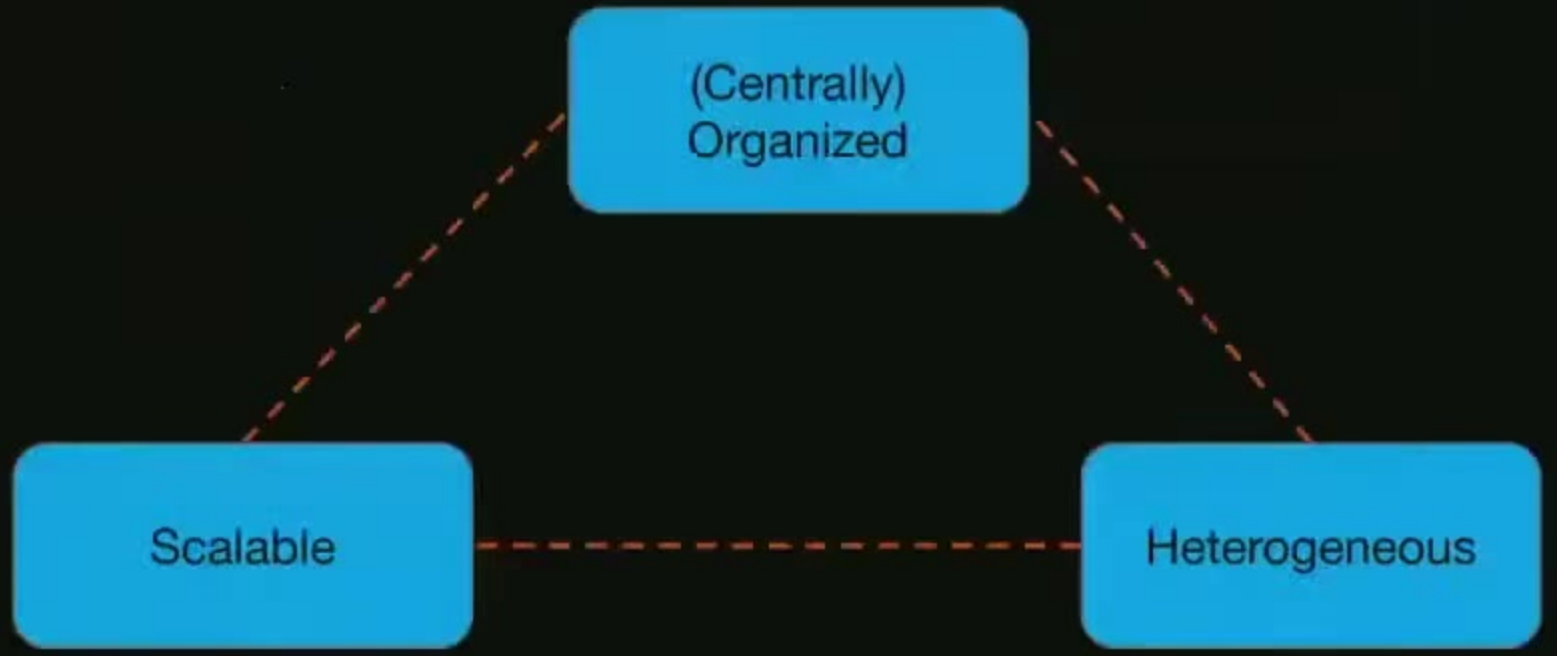
- Mitigate Bias
- Improve Robustness
- Model heterogeneity with MoE, Multimodal learning etc.

Without Heterogeneity

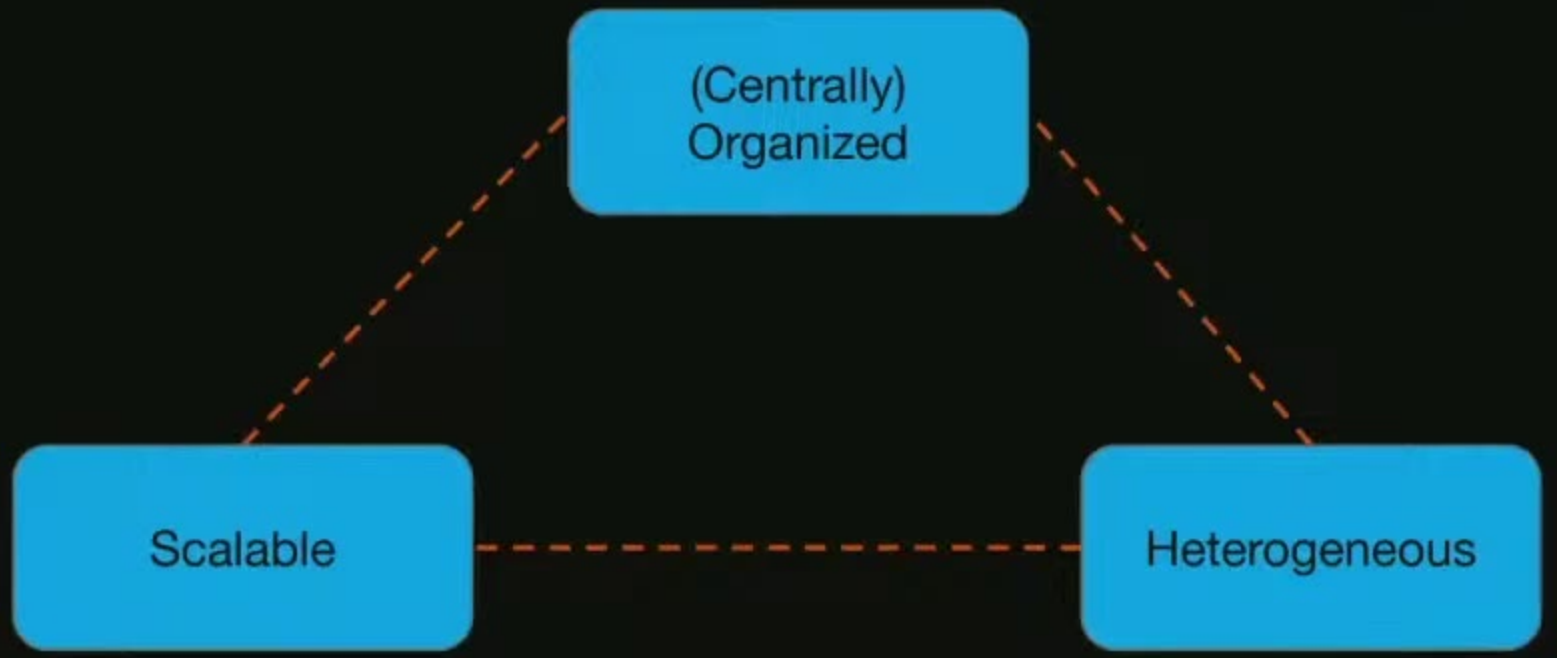
- Mode Collapse
- Neural Collapse
- Poor generalization

1. Bukharin, A., & Zhao, T. (2023). Data diversity matters for robust instruction tuning. arXiv preprint arXiv:2311.14736.
2. Brown et al., Language Models are Few-Shot Learners, NeurIPS'20

But there is a trilemma!



But there is a trilemma!



CoDream: Collaborative Dreams

Key Idea

- Based on DeepDream.
- Dreams capture wisdom stored inside neural network weights.
- We merge these dreams from multiple models and call it **CoDream**.
- Perform knowledge distillation with these dreams to train models.

Mordvintsev, Alexander; Olah, Christopher; Tyka, Mike (2015). "DeepDream - a code example for visualizing Neural Networks". Google Research

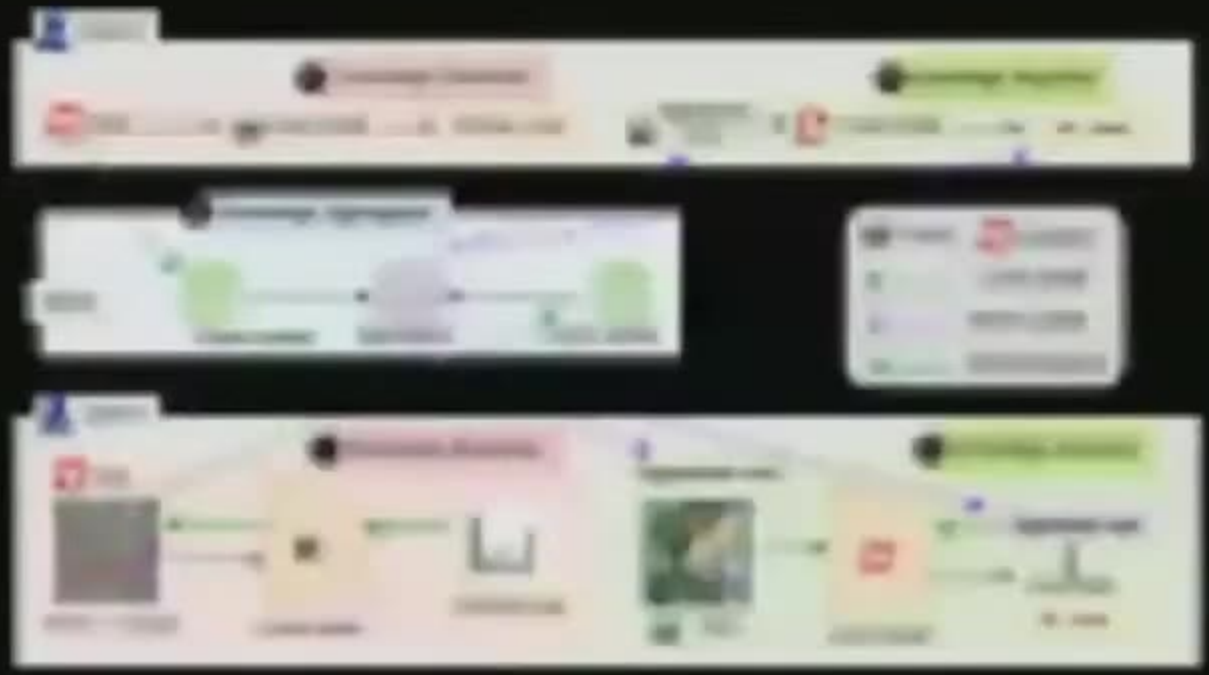
CoDream: Collaborative Dreams

Key Idea

- Based on DeepDream.
- Dreams capture wisdom stored inside neural network weights.
- We merge these dreams from multiple models and call it **CoDream**.
- Perform knowledge distillation with these dreams to train models.

Mordvintsev, Alexander; Olah, Christopher; Tyka, Mike (2015). "DeepDream - a code example for visualizing Neural Networks". Google Research

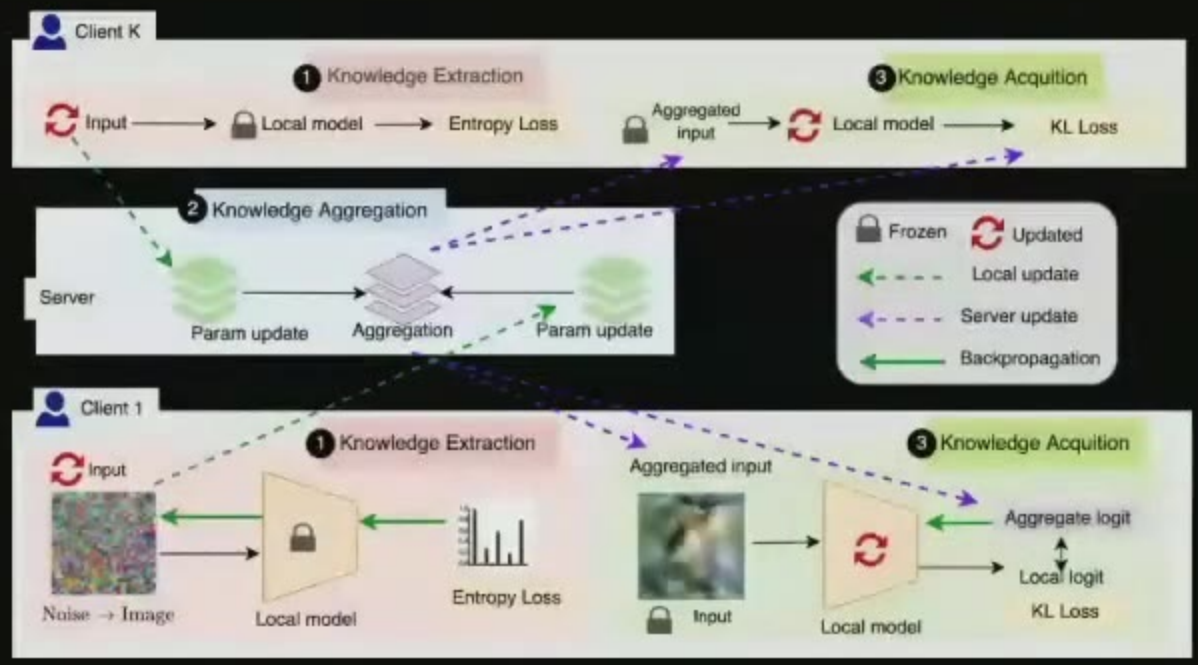
Co-Dream: Share representation not raw da



Algorithm

1. Start with a random image.
2. Optimize the random image locally.
3. Average these random images from multiple users.
4. Repeat 2. and 3. for ~1000 steps.
5. Optimize model weights on these CoDreams.

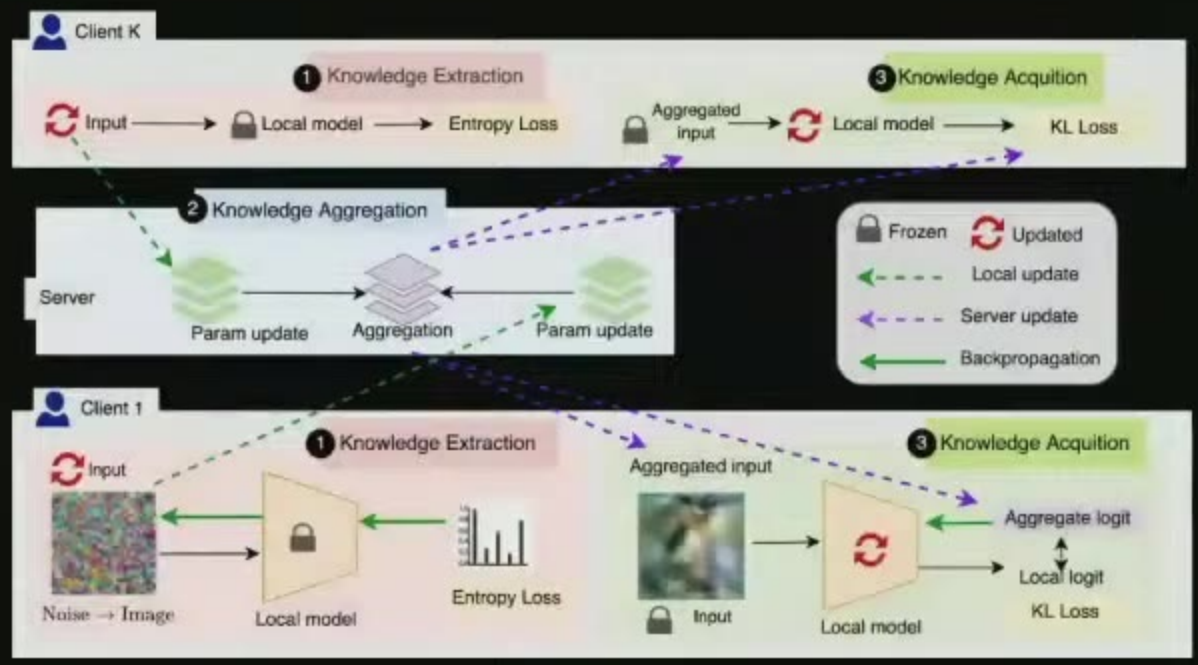
Co-Dream: Share representation not raw da



Algorithm

1. Start with a random image.
2. Optimize the random image locally.
3. Average these random images from multiple users.
4. Repeat 2. and 3. for ~1000 steps.
5. Optimize model weights on these CoDreams.

Co-Dream: Share representation not raw da



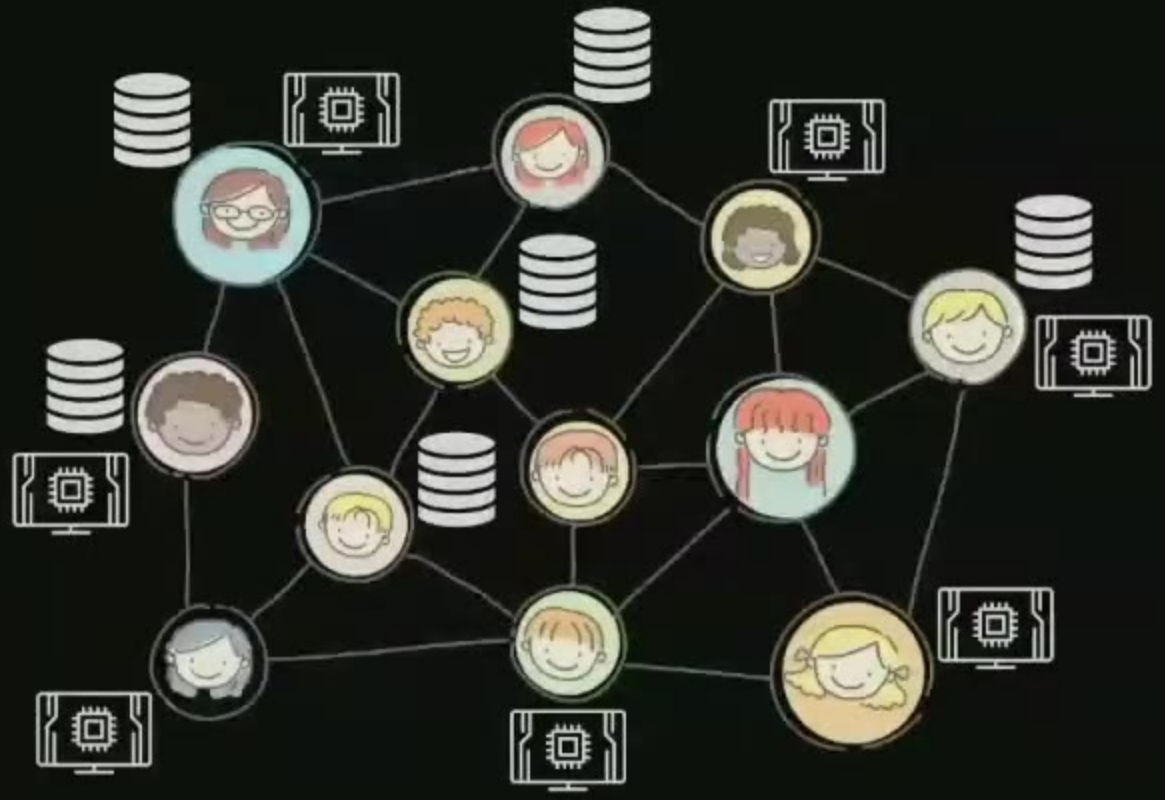
Algorithm

1. Start with a random image.
2. Optimize the random image locally.
3. Average these random images from multiple users.
4. Repeat 2. and 3. for ~1000 steps.
5. Optimize model weights on these CoDreams.

Co-Dream for scalable collaboration

- Not all users with computation resources have data
- Co-dream allows you to mine dreams from model parameters
  - No Data required!
  - Similar to proof of work (but this work is useful)
  - These dreams can be used to train new models.

Co-Dream for scalable collaboration

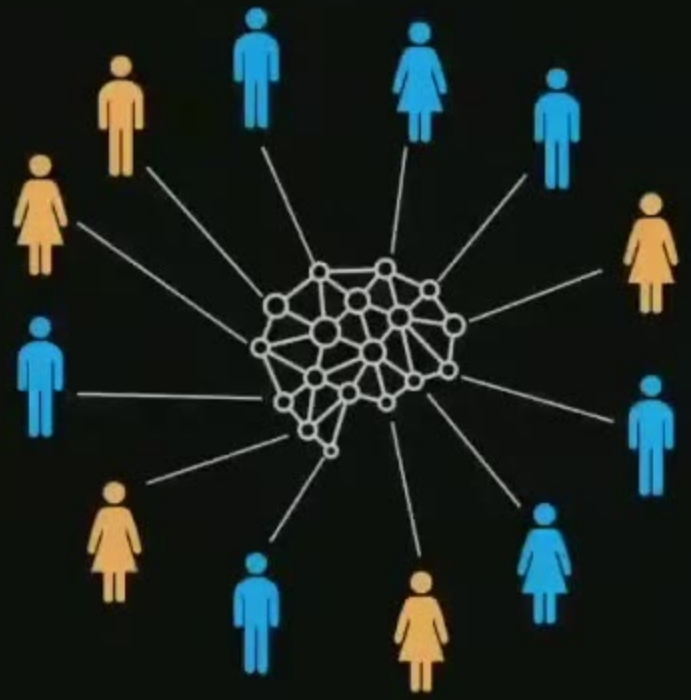


- Not all users with computation resources have data
- Co-dream allows you to mine dreams from model parameters
  - No Data required!!
  - Similar to proof of work (but this work is useful)
  - These dreams can be used to train new models.

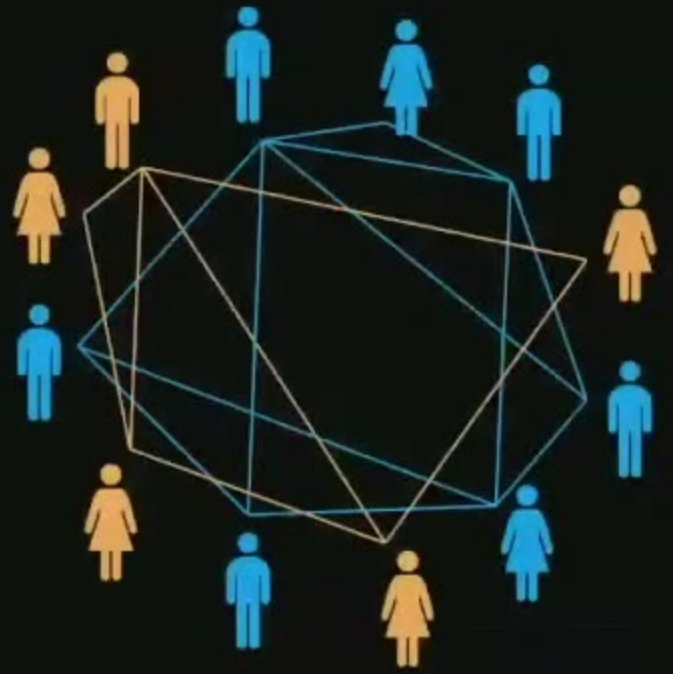
Co-Dream for scalable communication

Key Idea: Use Dreams to calculate similarity between data distributions and perform collaborator selection in a self-organized manner.

Homogeneous



Personalized



Putting it all together

1. Intelligence (Scalability and Heterogeneity) is the main goal.
2. CoDream - Synthesizing wisdom from raw data.
3. Heterogeneous Collaboration - Now users without any data can also contribute.
4. Collaborator Selection - Large scale collaborative learning without a central orchestrator.
5. Heterogeneous Models - Collaboration happens in the space of data not model parameters.

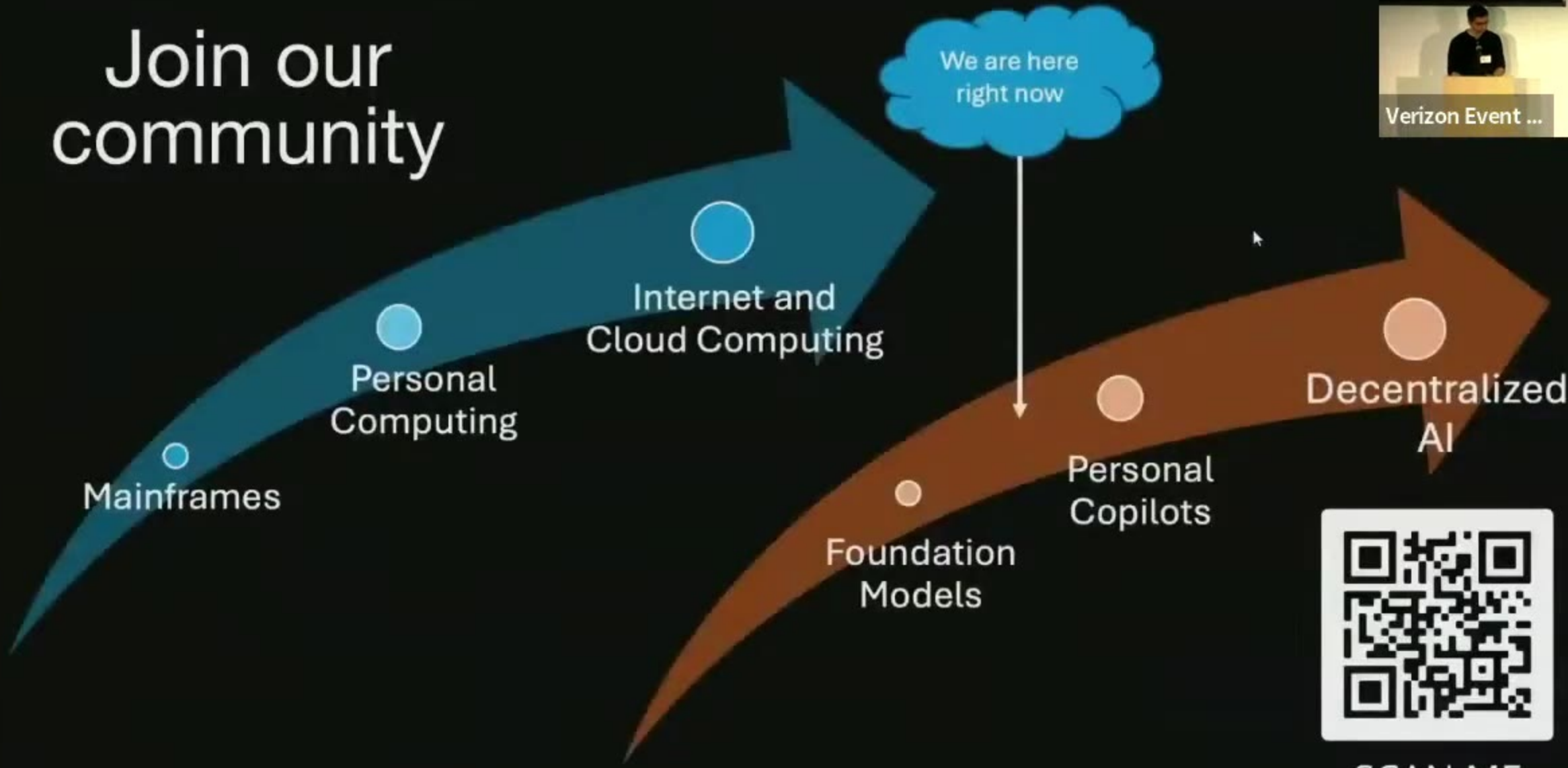
Putting it all together

1. Intelligence (Scalability and Heterogeneity) is the main goal.
2. CoDream - Synthesizing wisdom from raw data.
3. Heterogeneous Collaboration - Now users without any data can also contribute.
4. Collaborator Selection - Large scale collaborative learning without a central orchestrator.
5. Heterogeneous Models - Collaboration happens in the space of data not model parameters.

Putting it all together

1. Intelligence (Scalability and Heterogeneity) is the main goal.
2. CoDream - Synthesizing wisdom from raw data.
3. Heterogeneous Collaboration - Now users without any data can also contribute.
4. Collaborator Selection - Large scale collaborative learning without a central orchestrator.
5. Heterogeneous Models - Collaboration happens in the space of data not model parameters.

Join our community



SCAN ME