Low Latency Privacy Preserving Inference

Alon Brutzkus, Ran Gilad-Bachrach, Oren Elisha

Poster #176: Tuesday

https://github.com/microsoft/CryptoNets

Microsoft Research

Search

Home | My Network 1 | Jobs | Messaging | Notifi

Message | More...

Crook Ed 1st

Deep Inference Systems - Honestly (DISHONESTLY)

Azkaban University

Search

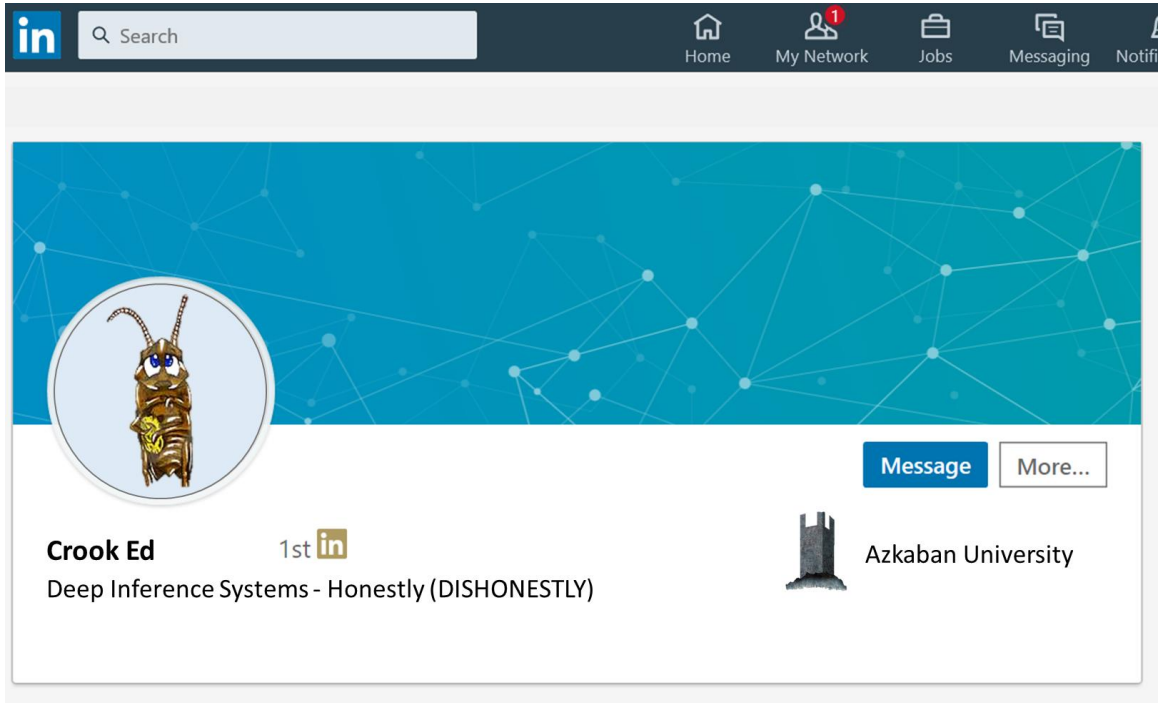
Home | My Network 1 | Jobs | Messaging | Notifi

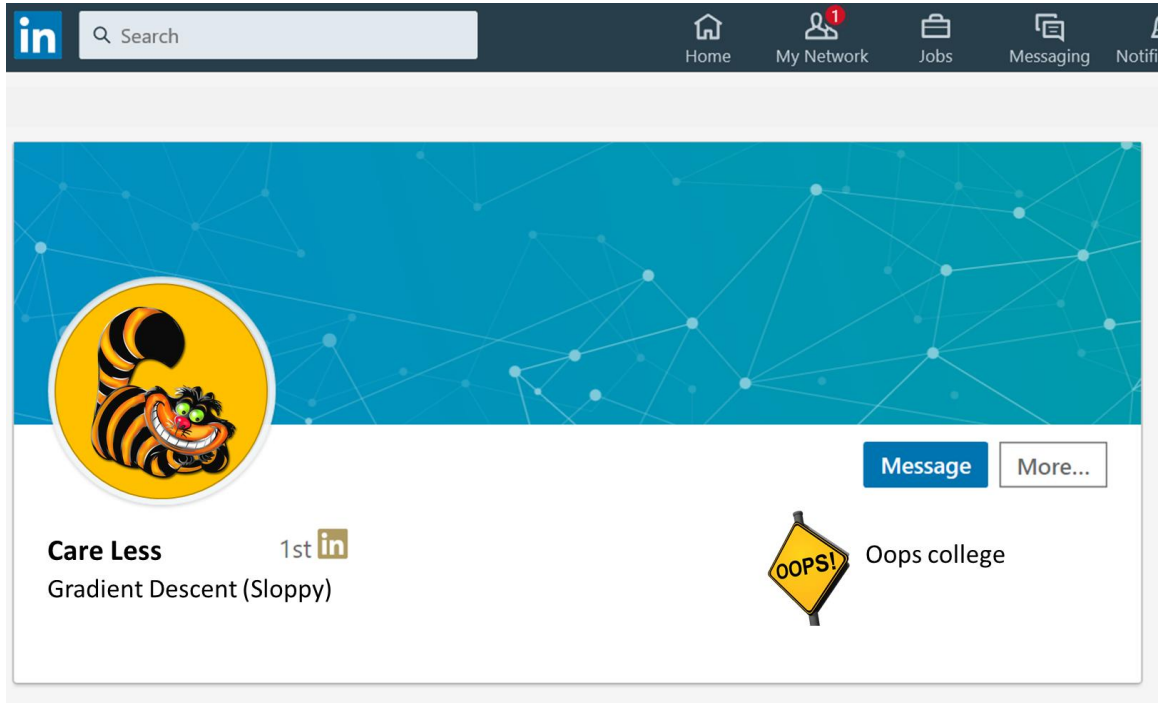
Message | More...

Care Less 1st

Gradient Descent (Sloppy)

Oops college





Prediction as a Service (PaaS):

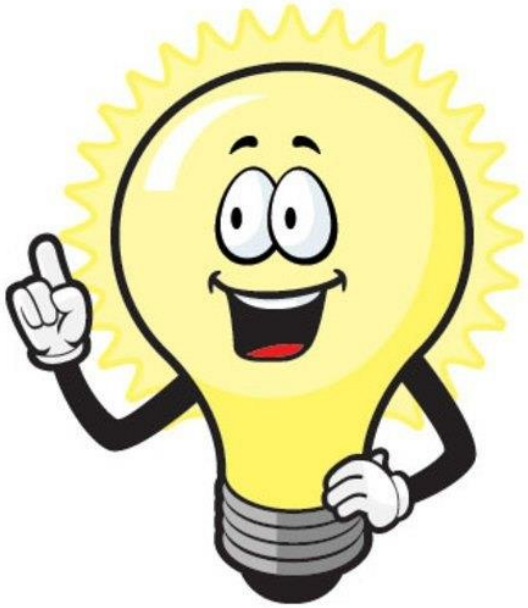
Send your private data (DNA, EMR, finance,…) and we'll send you back predictions

can Prediction as a Service be made trust worthy and efficient?

CryptoNets

1. The client encrypts the data and sends only the encrypted data to the service provider

2. The service provider applies the model to the encrypted data and sends the result



3. The client decrypts the message and reads the results

CryptoNets

1. The client encrypts the data and sends only the encrypted data to the service provider

2. The service provider applies the model to the encrypted data and sends the result

3. The client decrypts the message and reads the results

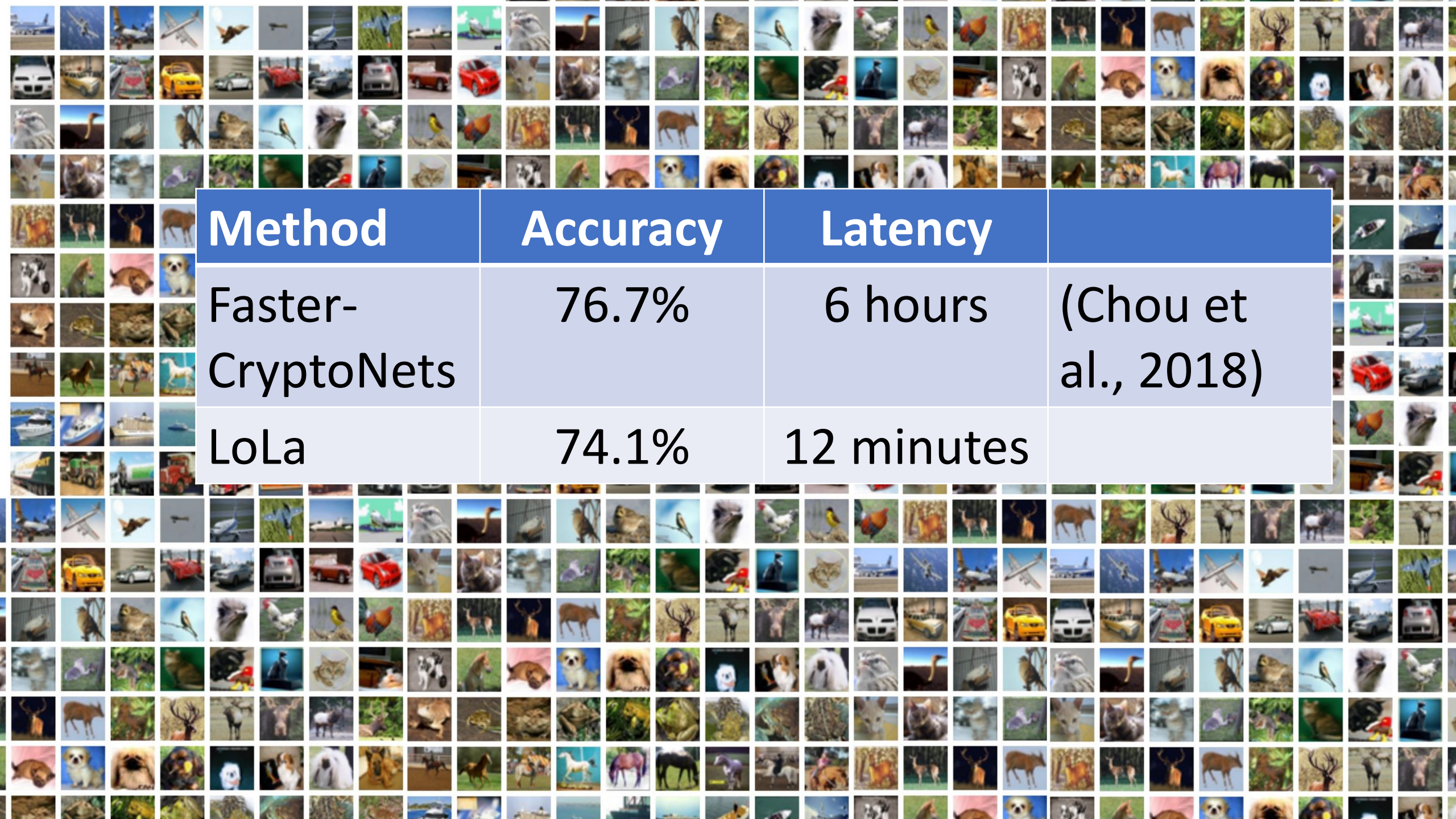
HOMOMORPHIC ENCRYPTION

HOMOMORPHIC ENCRYPTION

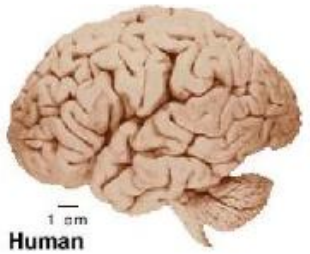
Method	Accuracy	Latency	
CryptoNets	98.95%	205s	(Dowlin et al., 2016)

Method	Accuracy	Latency	
CryptoNets	98.95%	205s	(Dowlin et al., 2016)
nGraph-HE	98.95%	135s	(Boemer et al., 2018)
Faster-CryptoNets	98.7%	39.1s	(Chou et al., 2018)
HCNN	99%	14.1s (GPU)	(Badawi et al., 2018)

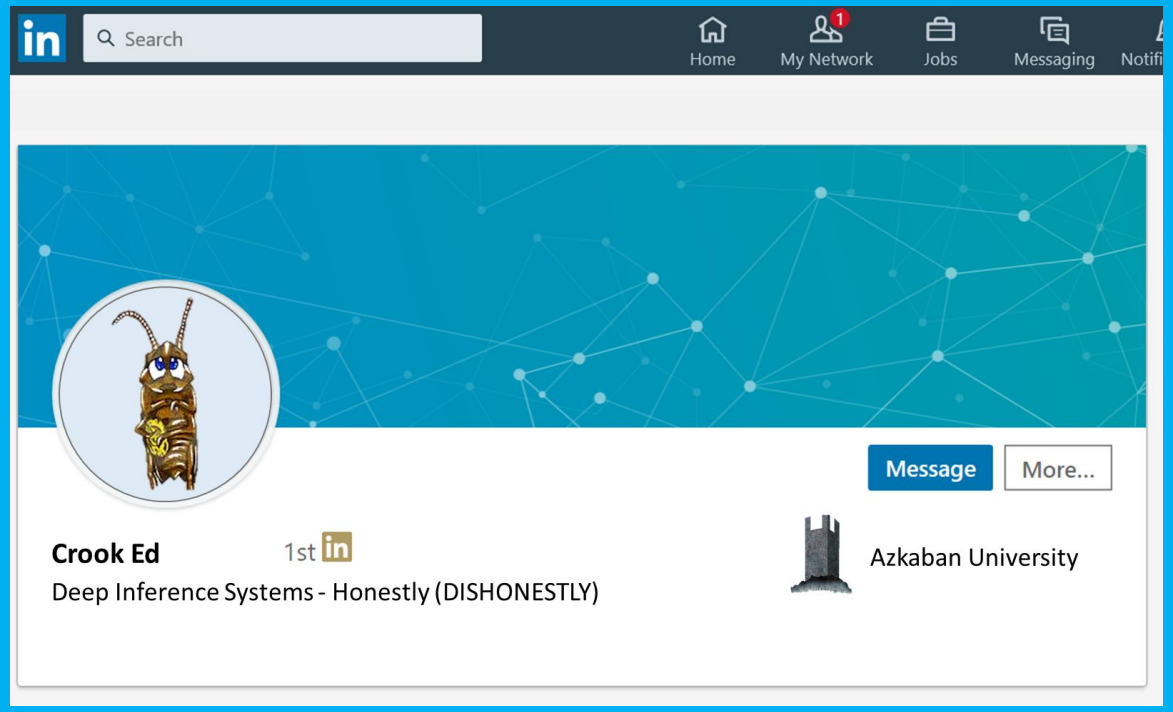
Method	Accuracy	Latency	
CryptoNets	98.95%	205s	(Dowlin et al., 2016)
nGraph-HE	98.95%	135s	(Boemer et al., 2018)
Faster-CryptoNets	98.7%	39.1s	(Chou et al., 2018)
HCNN	99%	14.1s (GPU)	(Badawi et al., 2018)
CryptoNets 2.3	98.95%	24.8s	
LoLa-Dense	98.95%	7.2s	
LoLa	98.95%	2.2s	

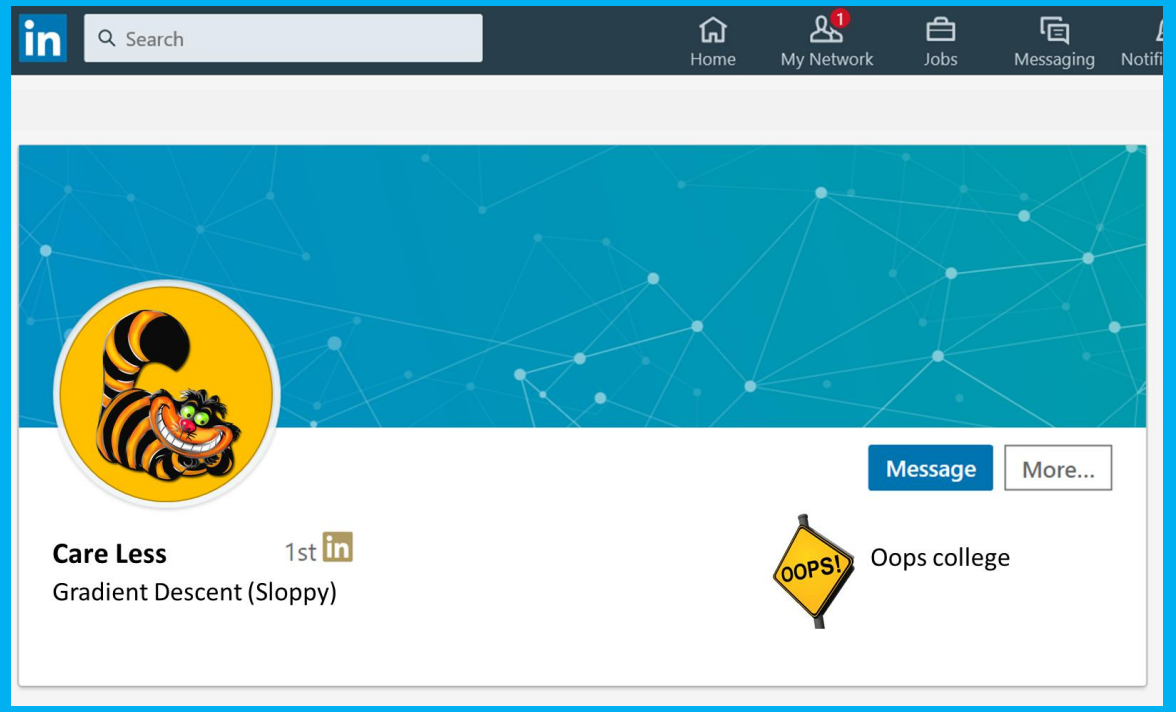


Method	Accuracy	Latency	
Faster-CryptoNets	76.7%	6 hours	(Chou et al., 2018)
LoLa	74.1%	12 minutes	

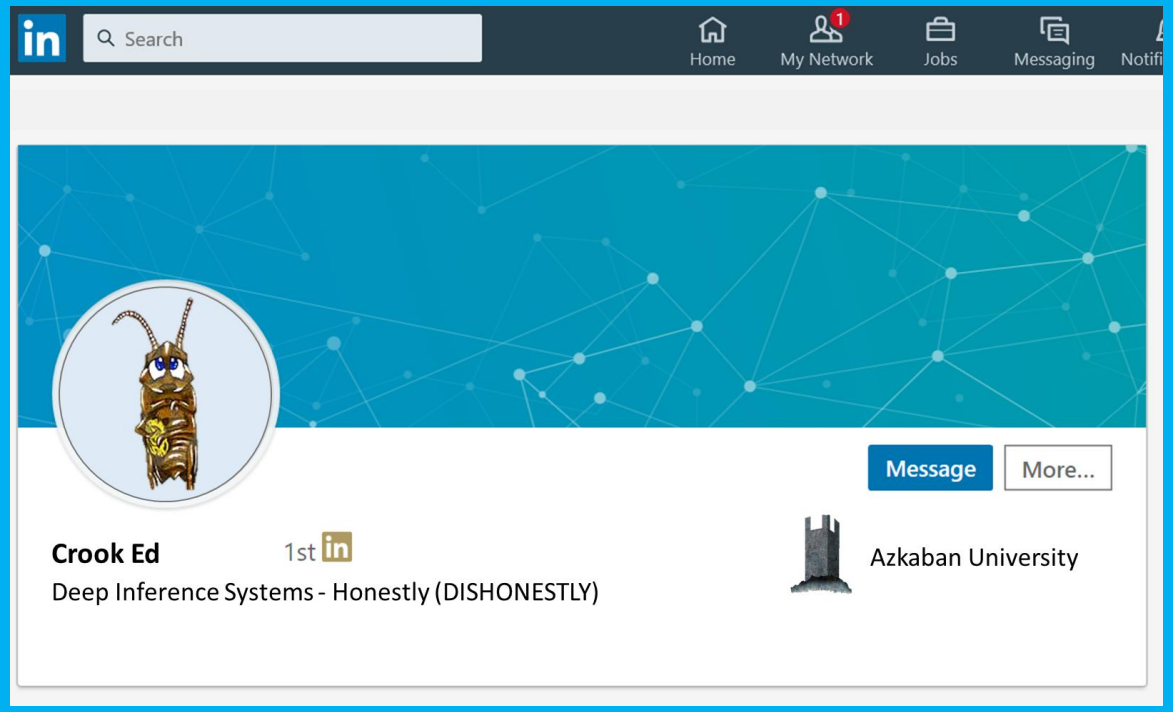


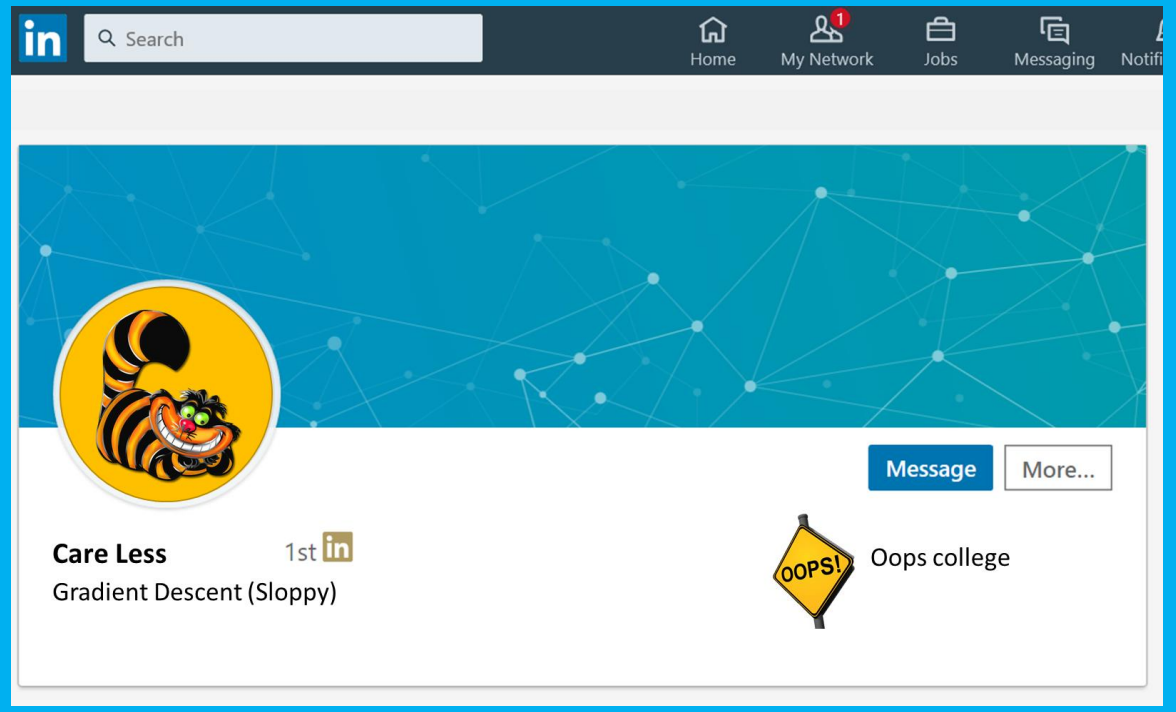
Method	Accuracy	Latency	
LoLa-AlexNet	81.6s	0.16s	





can Prediction as a Service be made trust worthy and efficient?





can Prediction as a Service be made trust worthy and efficient?

YES WE CAN

How?

Seriously?

Training?

Low Latency Privacy Preserving Inference

Alon Brutzkus, Ran Gilad-Bachrach, Oren Elisha

Poster #176: Tuesday

https://github.com/microsoft/CryptoNets

Microsoft Research