Breaking Down AI to get Explanations

Karel Hynek
AI is everywhere
Me!

- **PhD Candidate** at FIT Czech Technical University
- Four years developing **Network security detectors**
- Mainly focusing on **AI-based security detectors**
Accuracy of 99.9%
But it is hard to verify the output.
Are those hotdogs?

YES!
Is this malware?

Packet Sizes: [231|412|221|82|232|415|240|82|242|414|230|82|237|415|244|82|261|75|412|64],

Interpacket Times: [10 | 14 | 443|13|14 | 53 |534|44|424|232|535|44|232 | 1 | 13 |44|434|32|555|23]

Directions: [1 | -1 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | 1 | 1]
Packet Sizes: [231|412|221|82|232|415|240|82|242|414|230|82|237|415|244|82|261|75|412|64]
Interpacket Times: [10 | 14 |443|13|14 | 53|534|44|424|232|535|44|232 | 1 |13 |44|434|32|555|23]
Directions: [1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | 1 |-1 | 1 | 1]
Is this malware?

Packet Sizes: [231, 412, 221, 82, 232, 415, 240, 82, 242, 414, 230, 82, 237, 415, 244, 82, 261, 75, 412, 64]


Directions: [1, -1, 1, 1, 1, 1, -1, -1, 1, -1, 1, 1, -1, -1, 1, -1, 1, -1, 1, -1, 1, 1]
Quick Math...

10K Network telemetry records per second

99.9% Accuracy

1 false-detection per second

Network of mid-sized company

Higher standard for network threat classification

10K * 0.001

It is gonna be turned off very soon!
We need to distinguish wrong detections quickly... By adding explanations.
Monolithic AI
Dividing AI

Because there is Bun and sausage
Is this a hotdog?

YES!
Is this a hotdog?

YES!
Is this a hotdog?
Is this a hotdog?

Yes!

Something is really wrong!
Is this a hotdog?

Yes!

Something is really wrong!
Dividing AI

Periodic Communication

Covert Communication

Crypto Mining

P2P communication

Because I've found the following properties...

YES!
Is this communication malicious?

Packet Sizes: [231|412|221|82|232|415|240|82|242|414|230|82|237|415|244|82|261|75|412|64]

Interpacket Times: [10 | 14 | 443|13|14 | 53 |534|44|424|232|535|44|232 | 1 | 13 |44|434|32|555|23]

Directions: [1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | 1 | -1 | 1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1]
The benefits of divided AI

1. We can quickly identify misclassifications
2. Less prone to design errors
3. Even non-AI expert can understand the explanations
You can contact me at karel.hynek@cesnet.cz