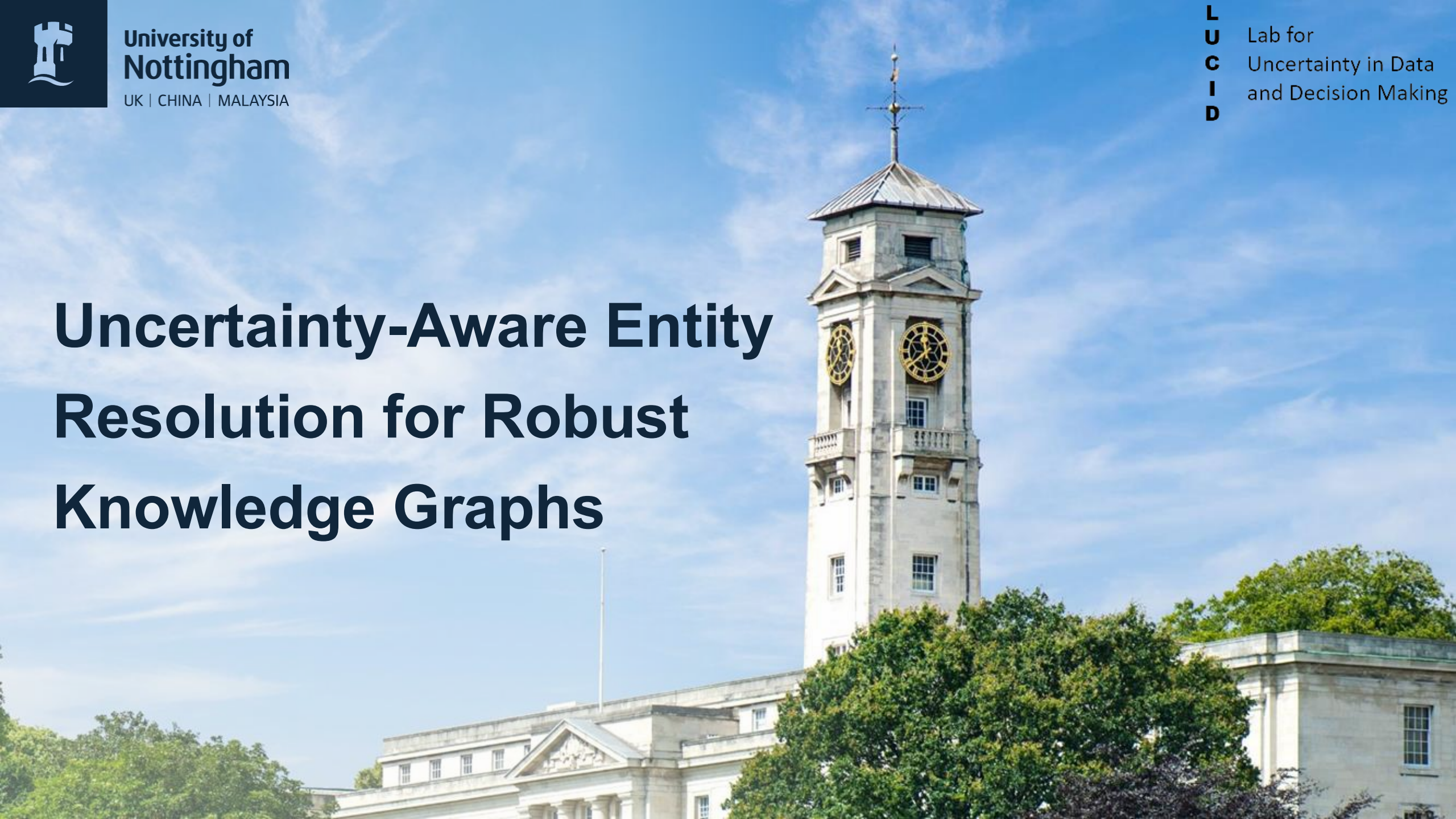


Uncertainty-Aware Entity Resolution for Robust Knowledge Graphs

Outline

- Introduction
- The Project
- Methodologies
- Future Work

Introduction

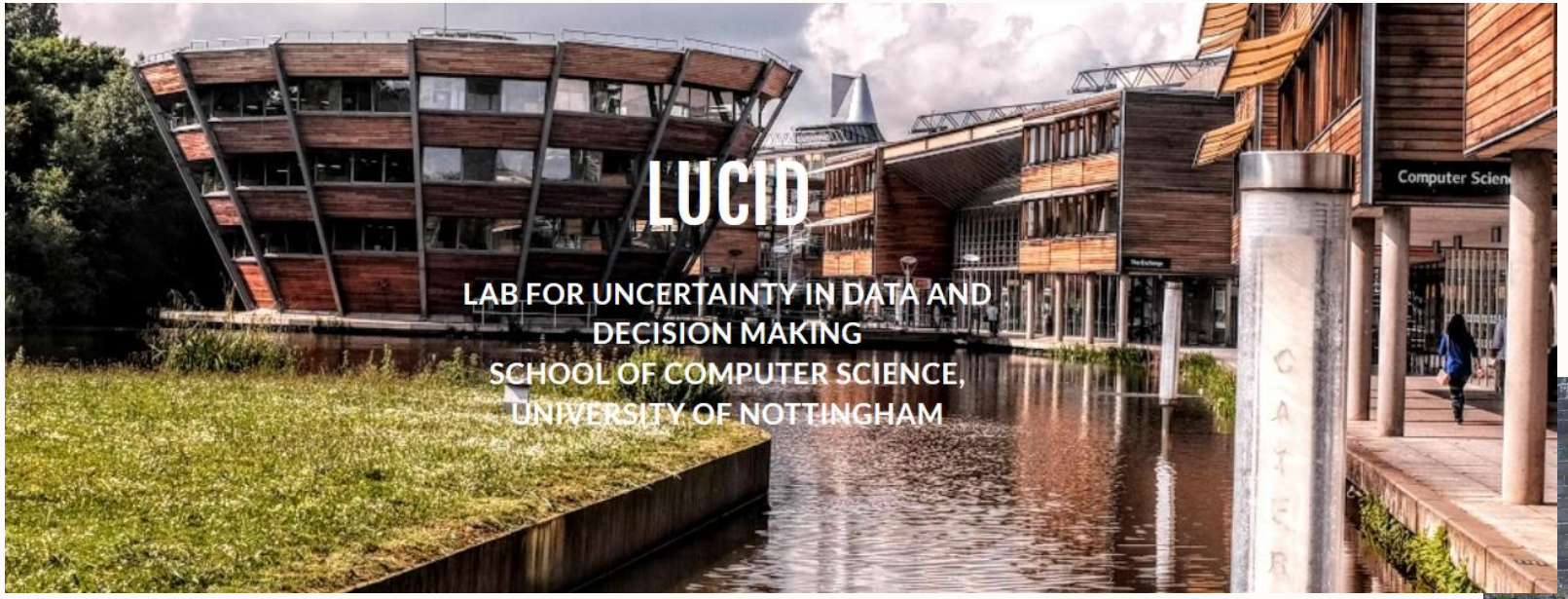
University of Nottingham:

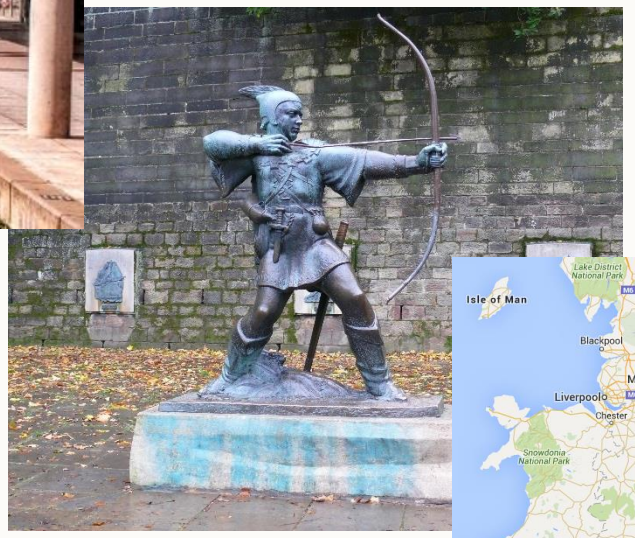
- Direnc Pekaslan, Assistant Prof, PI
- Christian Wagner, Prof., Co-I
- Nasser Alkhulaifi, Researcher

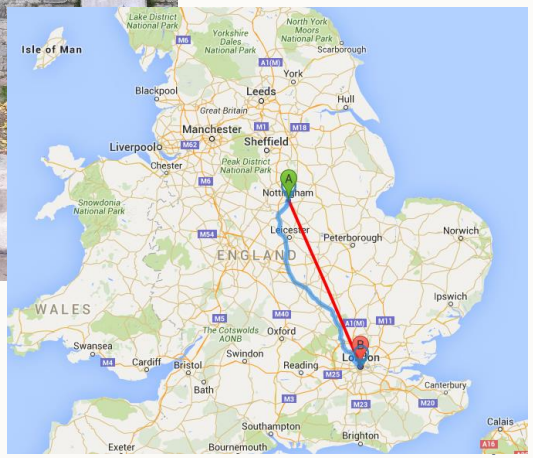
The Alan Turing Institute

- Agnė Šniukštaitė-Clemente - Research Project Manager
- Nigel D.

Introduction







- LUCID (http://lucidresearch.org)
  - Lab for Uncertainty in Data and Decision Making

- Based at the School of Computer Science, University of Nottingham, UK

What we do…

- End-to-end handling of uncertainty, from information capture, to modelling, inference and communication – *for better-informed decision making*.

- For example, uncertainty handling in:
  - Cybersecurity vulnerability assessment
  - Environmental management
  - FMCG / product development

The Project

The Project

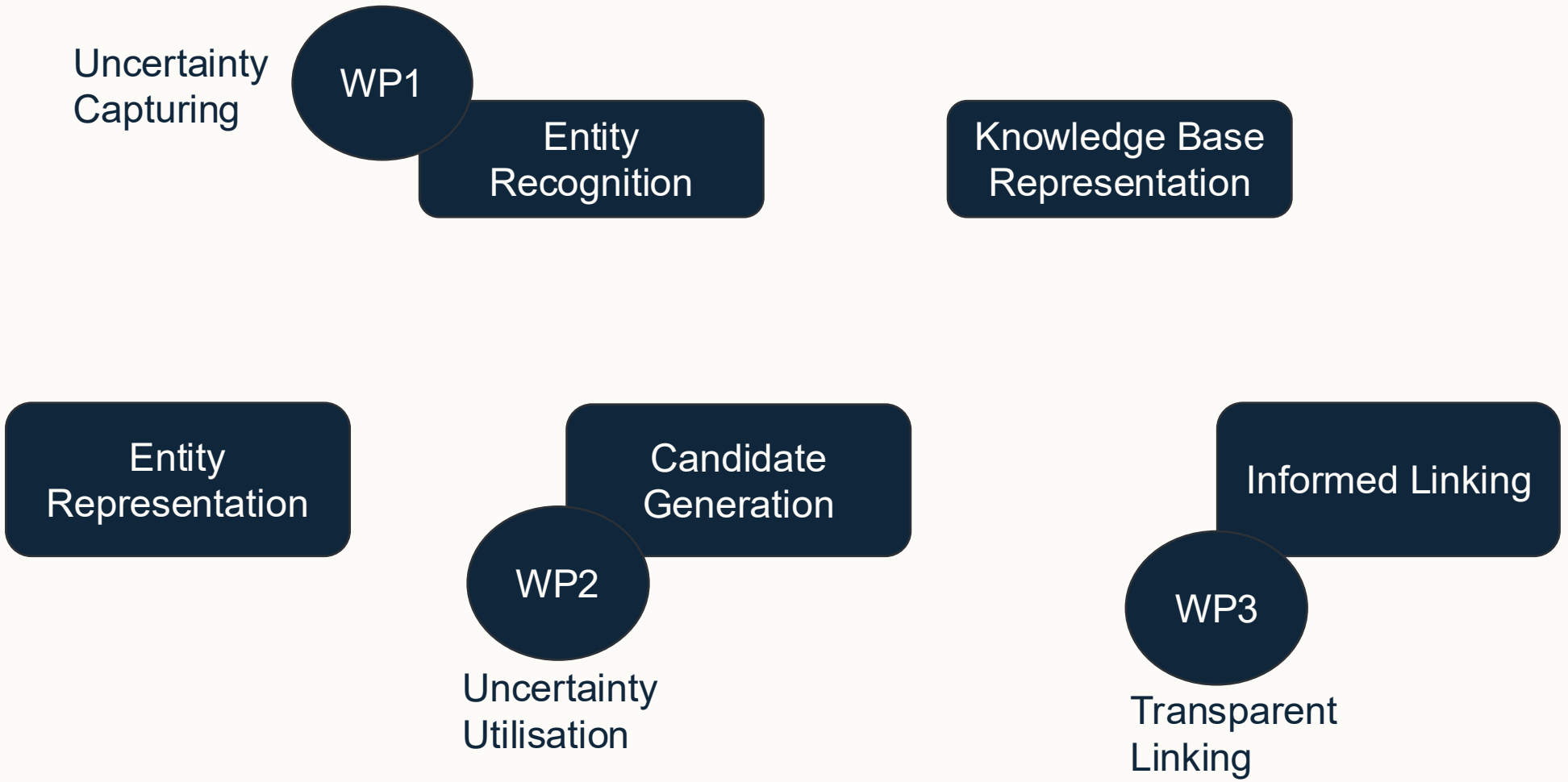
- Our goal is to understand and show the utility of handling uncertainty across the entity resolution pipeline
  - from capturing different aspects of uncertainty, to modelling and communicating it.

- As part of the initial project, we selected a few methodologies to act as ‘probes’.

- We focused on
  - *span boundary uncertainty* and *type uncertainty* of entities via model instability

- Challenges:
  - Evaluation vs state-of-the-art results
  - Establishing utility: ‘How useful is it in practice? Does it result in better-informed decision-making?’

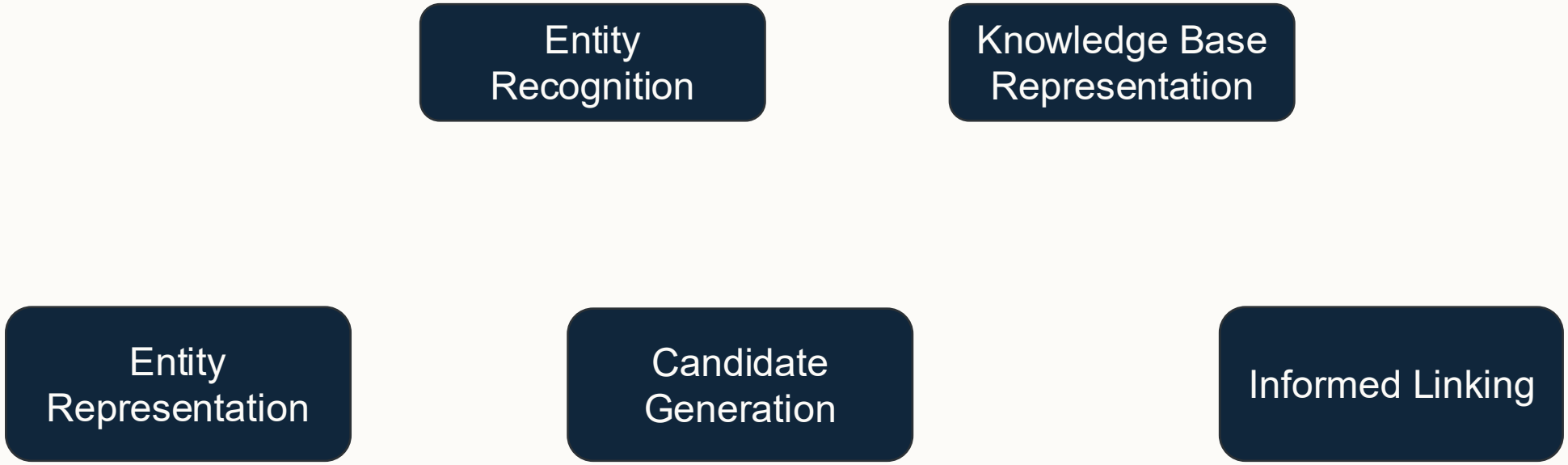
Project Overview - Entity Resolution Pipeline



Pilot Methodologies

* Probabilistic Weighted Span with soft Type Filtering
* Uncertain Embeddings with pyNNDescent

Project Overview - Entity Resolution Pipeline



Dataset

AnnoCTR Source dataset:

Cybersecurity-Specific Layer - 120 Reports

Mentions : 13,244

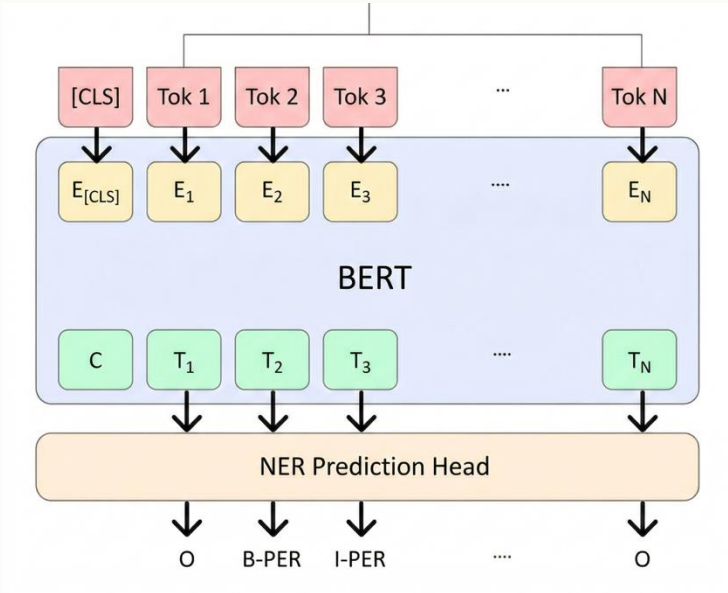
Types: LOC, ORG, SECTOR, TOOL, MALWARE, GROUP

60% Training, 15% validation and 25% Testing

Entity Recognition - BERT

BERT (Bidirectional Encoder Representations from Transformers)

“Squirrelwaffle : New Loader Delivering Cobalt Strike”

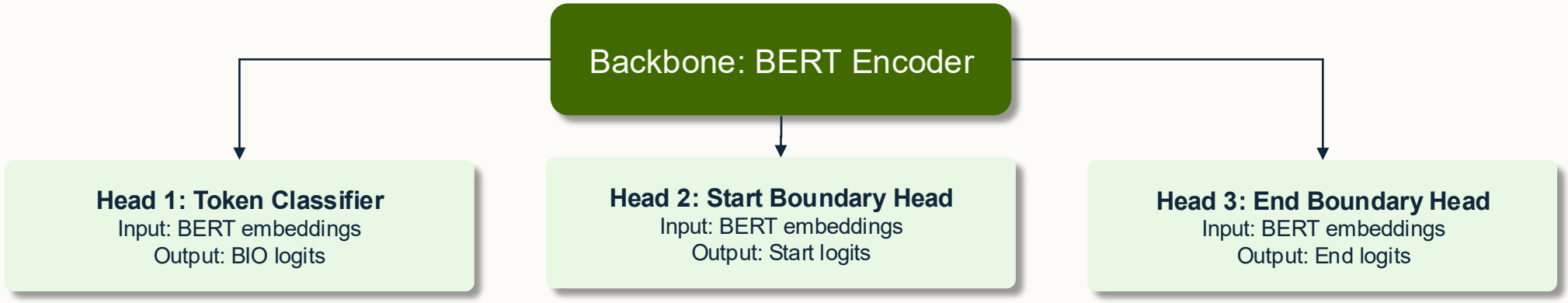


“Squirrelwaffle : New Loader Delivering Cobalt Strike”

O O O O B-MAL I-MAL

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019, June. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Type & Span Detection

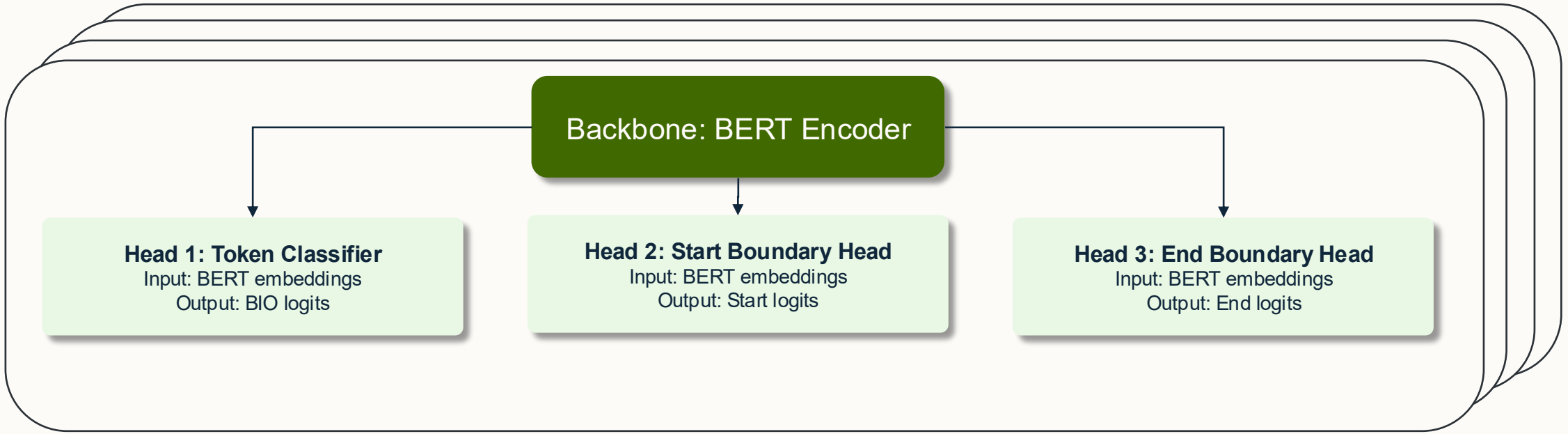


Model Architecture Rationale:
1. **What** entities are/types (token classification)
2. **Where** entities start and end (boundary detection)

Capturing Uncertainty

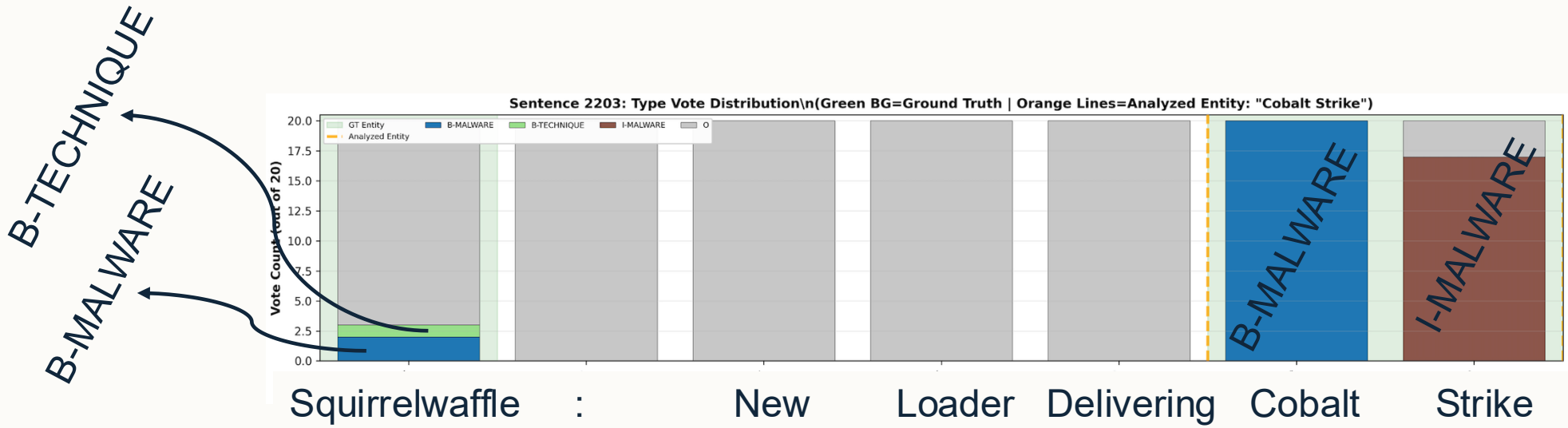
Type & Span Uncertainty ‘Generation’

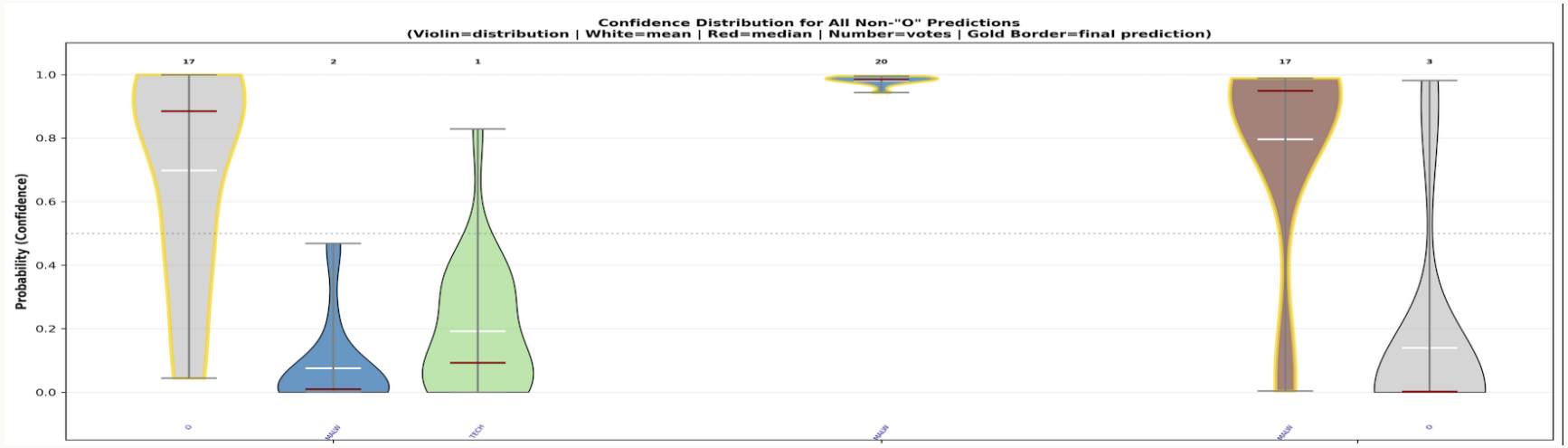
Entity Recognition



Limitation: MC-dropout focuses on model-uncertainty…

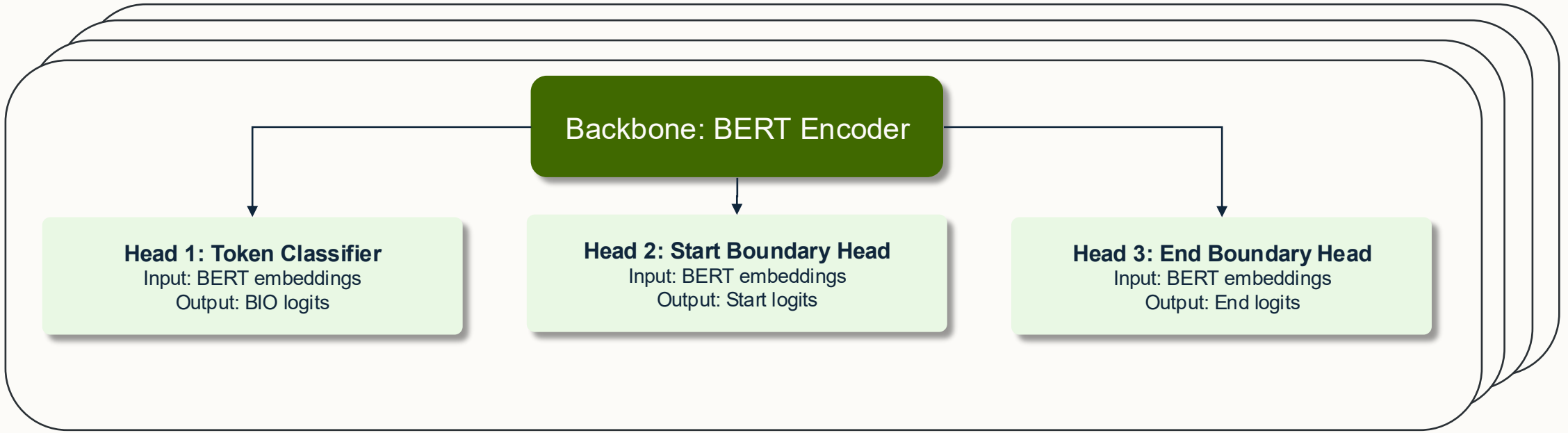
Type Uncertainty – via Head 1





Type & Span Uncertainty ‘Generation’

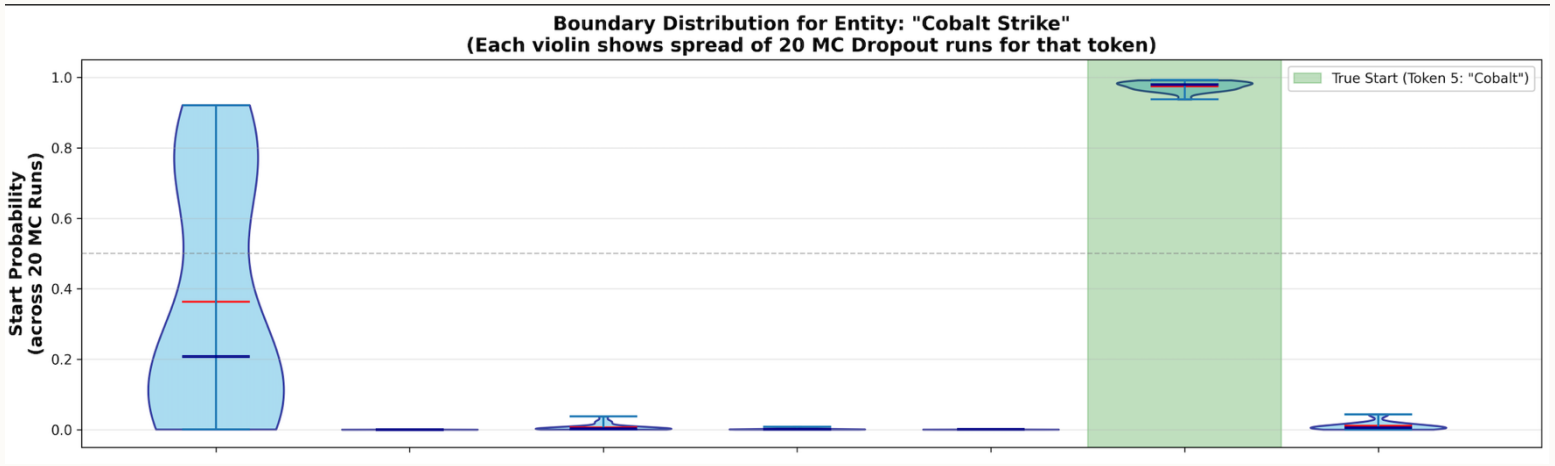
Entity Recognition



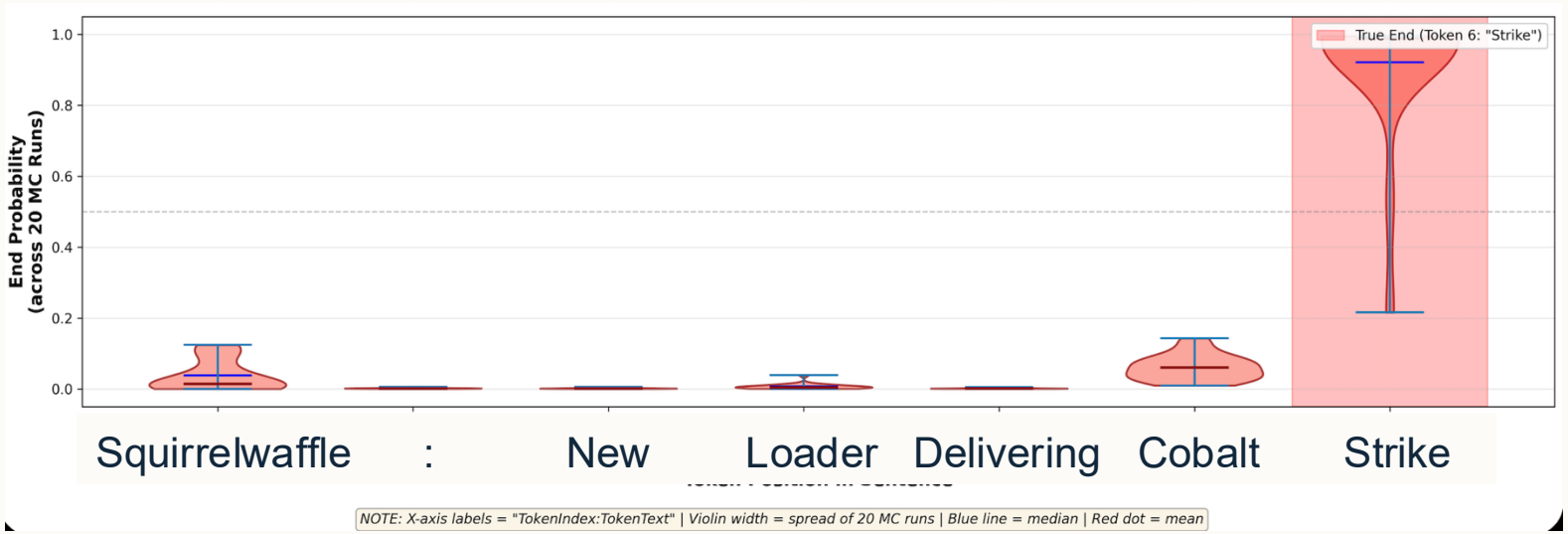
Span Uncertainty - via Heads 2 & 3

BERT - MC Dropout – Heads 2 & 3 Span Detection

Span Start



Span End



BERT vs BERT MC Dropout

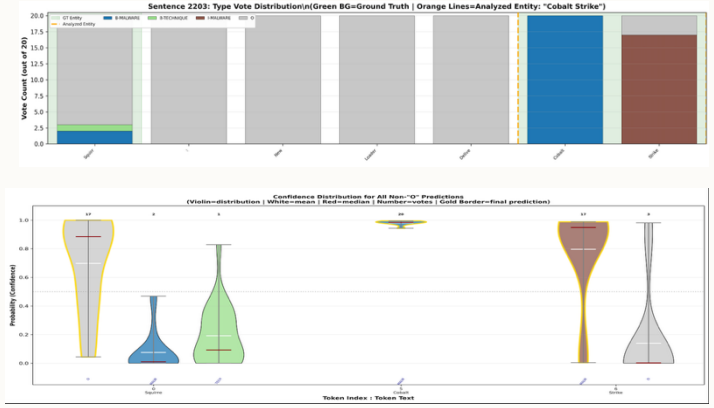
BERT

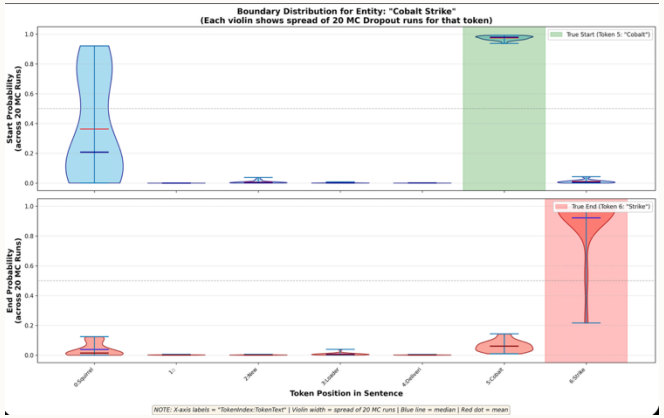
Squirrelwaffle : New Loader Delivering Cobalt Strike

O O O O B-MAL I-MAL

BERT - MC Dropout

Squirrelwaffle : New Loader Delivering Cobalt Strike

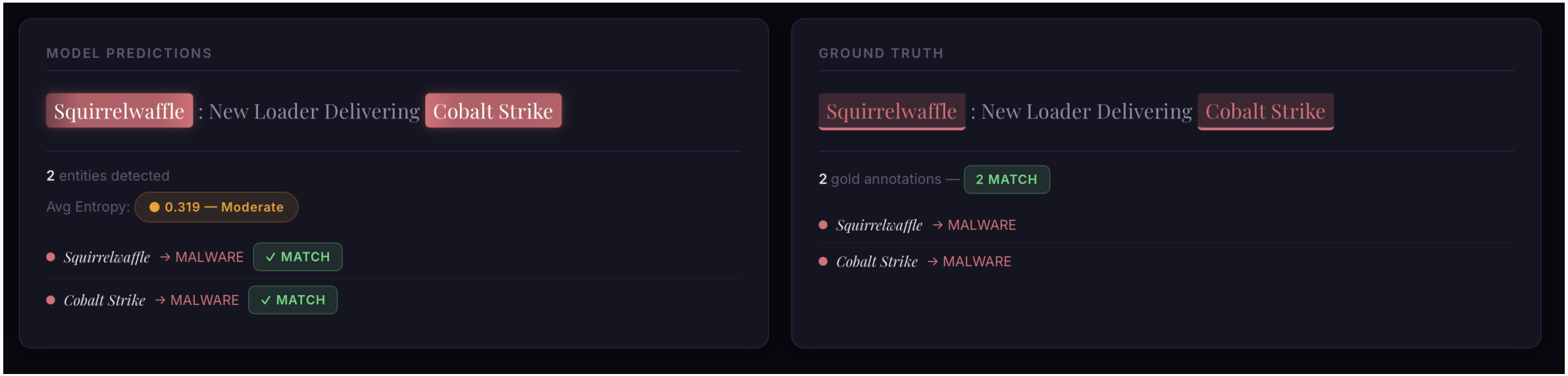




- Type Uncertainty
- Span Uncertainty

Dashboard Demo

Entity Recognition



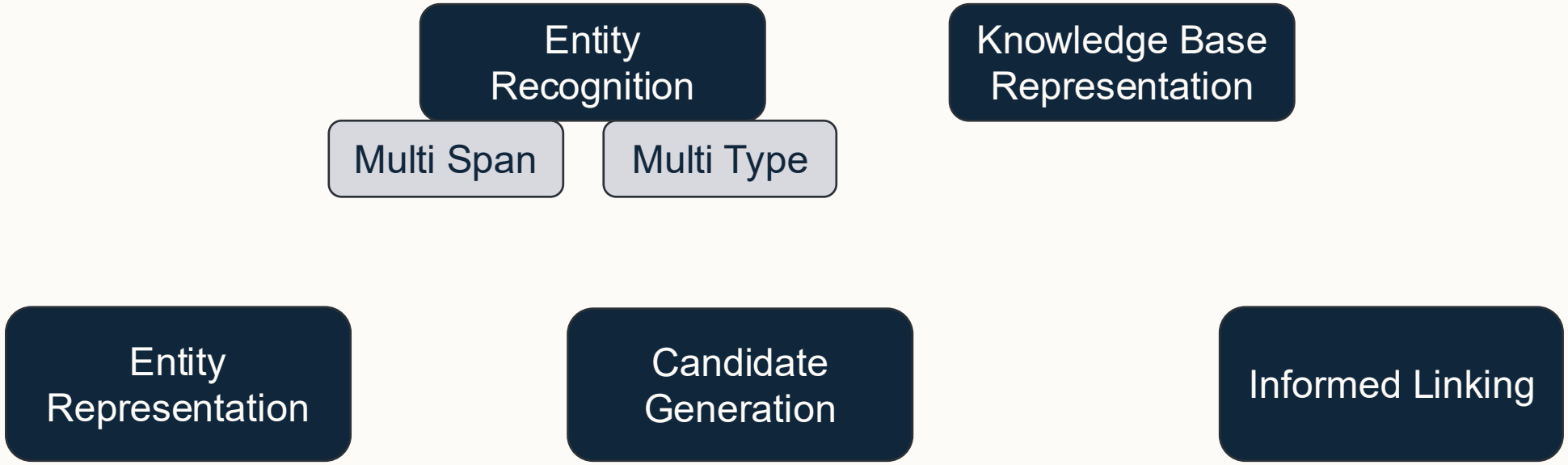
Dashboard Demo

Entity Recognition

Modelling Uncertainty

Going beyond uncertainty capture via MC dropout

Project Overview - Entity Resolution Pipeline



Modelling Span Boundary Uncertainty

Current Approach:

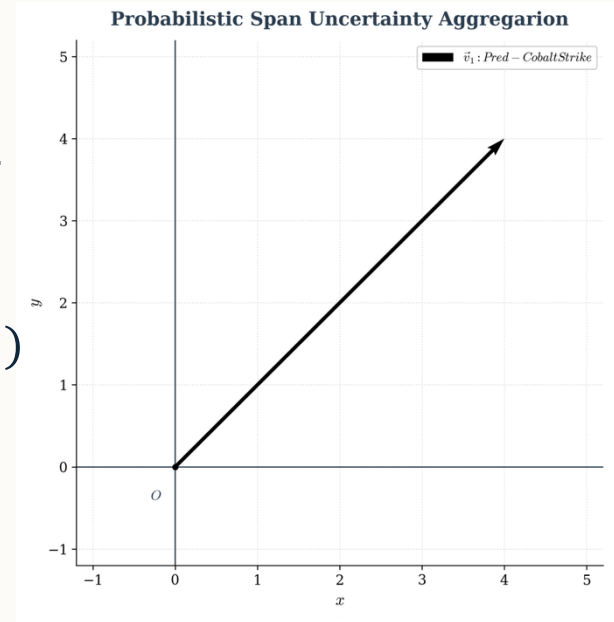
Multi Span

{Cobalt, Cobalt Strike}

FastText

$$\text{embed}_{\text{idf}}(s_k) = \frac{\sum_{t \in T_k} w_t^{\text{idf}} \cdot \mathbf{e_t}}{\sum_{t \in T_k} w_t^{\text{idf}}}$$

$$\mathbf{v_{mc}} = \sum_{k=1}^{K} P(s_k) \cdot \text{embed}_{IDF}(s_k)$$



- IDF
- Probabilistic Span

Modelling Span Uncertainty +

Entity Representation

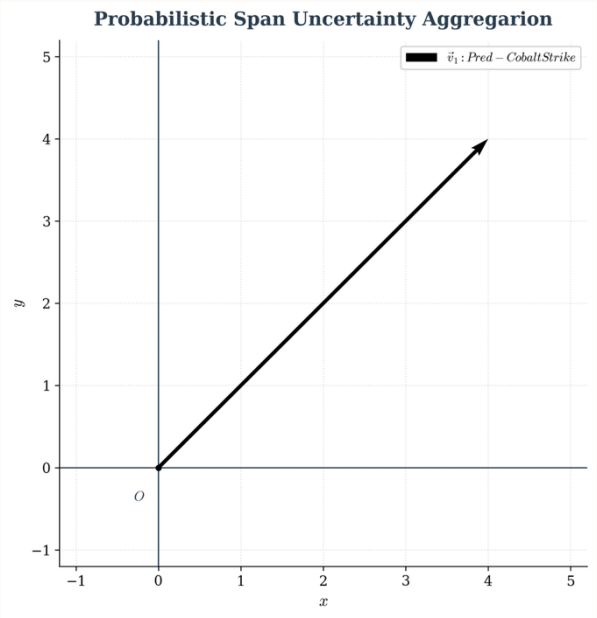
Multi Span

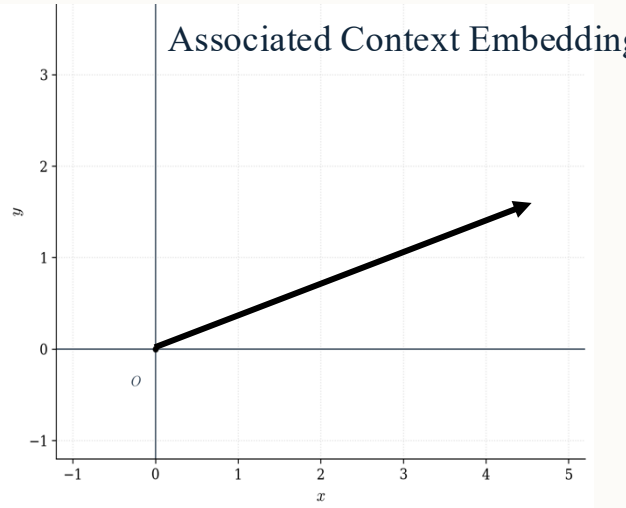
{Cobalt, Cobalt Strike}

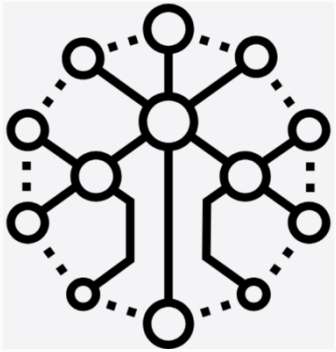
{Loader Delivering Cobalt Strike}

{Squirrelwaffle, Cobalt Strike}

FastText Dictionary







Entity Embedding
Capturing Span Uncertainty

Associated Window
of the Entity

Associated Neighbours

Project Overview - Entity Resolution Pipeline

Entity Recognition

Multi Span

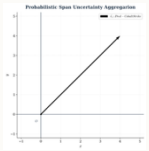
Multi Type

Knowledge Base Representation

Entity Representation

Candidate Generation

Informed Linking



Knowledge Base Representation

Towards *linking* to a KB

Knowledge Base Dataset

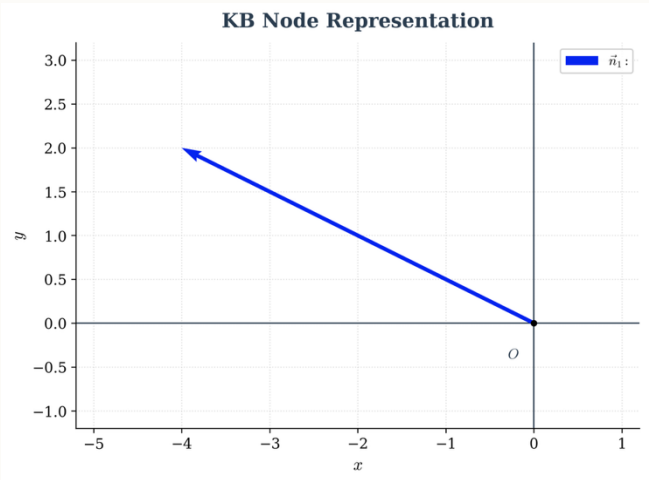
MITRE ATT&CK : Accessible knowledge base of adversary behaviours based on real-world observations.

AnnoCTR Entity Type	MITRE ATT&CK Object	Description	Real-World Example Span	Example Linked ID
TACTIC	Tactic (TA-)	The high-level objective or goal the adversary is trying to achieve.	"establish persistent access"	TA0003 (Persistence)
TECHNIQUE	Technique/Subtechnique (T-)	The specific method or action used to achieve a tactic.	"dump credentials from the system memory"	T1003.001 (OS Credential Dumping: LSASS Memory)
GROUP	Group (G-)	Tracked threat actors, hacker groups, or APTs.	"Lazarus Group"	G0032 (Lazarus Group)
MALWARE	Software: Malware (S-)	Malicious payloads, viruses, or ransomware.	"Squirrelwaffle"	S1030 (Squirrelwaffle)
TOOL	Software: Tool (S-)	Commercial, open-source, or built-in utilities used maliciously.	"Mimikatz"	S0002 (Mimikatz)

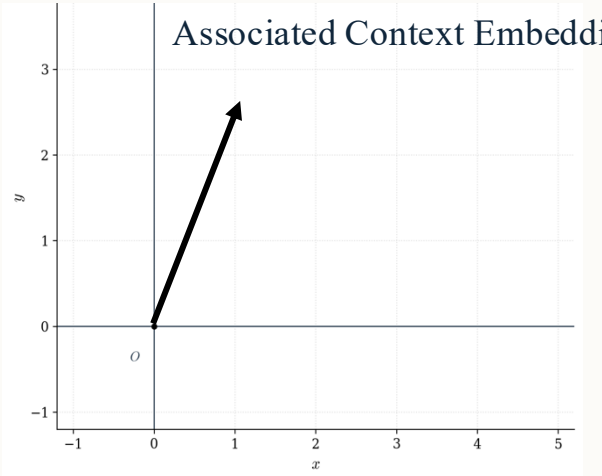
Knowledge Base Dataset

MITRE ATT&CK : Accessible knowledge base of adversary behaviours based on real-world observations.

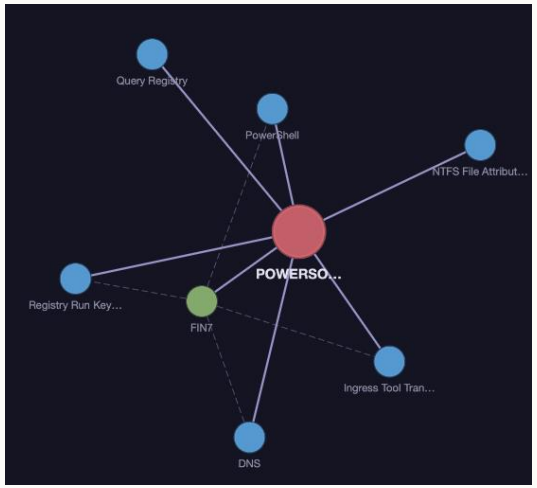
"**kb_type**": "malware",
"**name**": "POWERSOURCE",
"**aliases**": ["DNSMessenger"],
"**description**": "[POWERSOURCE] is a PowerShell backdoor that is a heavily obfuscated….



Name and Aliases



Description



Associated Neighbours

Knowledge Base Dataset

Knowledge Base Representation

S0145 POWERSOURCE MALWARE

ALIASES:

DNSMessenger

field_a: powersource dnsmessenger

CONTEXTUAL DESCRIPTION (FIELD_B)

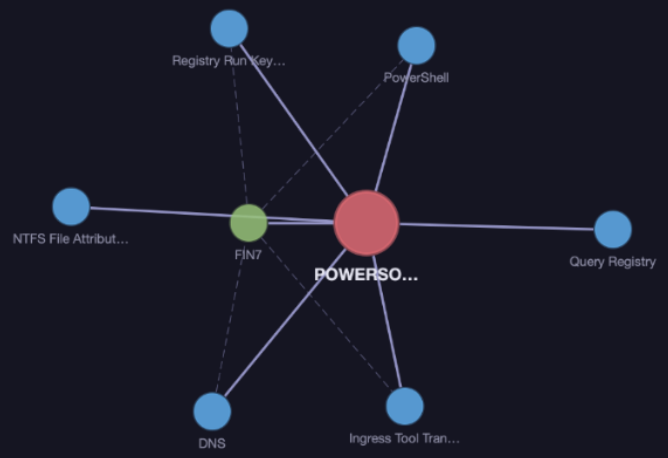
RIPPLE — CONTEXT EXPANDING FROM ENTITY CORE

powersource https attack mitre org software s0145 is a powershell backdoor that is a heavily obfuscated and modified version of the publicly available tool dns txt pwnage it was observed in february 2017 in spearphishing campaigns against personnel involved with united states securities and exchange commission sec filings at various organizations the malware was delivered when macros were enabled by the victim and a vbs script was dropped citation fireeye fin7 march 2017 citation cisco dnsmessenger march 2017

KNOWLEDGE GRAPH NEIGHBOURHOOD

7 connected entities — drag nodes to rearrange, scroll to zoom

Group Malware Technique



T1547.001 Registry Run Keys / Startup Folder TECHNIQUE

Candidate Generation

Measures to consider

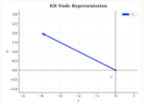
Project Overview - Entity Resolution Pipeline

Entity Recognition

Multi Span

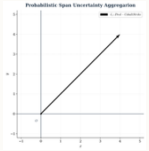
Multi Type

Knowledge Base Representation



Name | Neighbours

Entity Representation



Entity | Neighbours

Candidate Generation

Informed Linking

Softy Type Filtering / Compatibility

Candidate Generation

- Inverted Index Lookup $\mathrm{IDF}(t) = \log \frac{N}{|\{d : t \in d\}|}$

- Lexical Score $\mathrm{score\_lex}(m, c) = \frac{|Q \cap K(c)|}{|Q|}$

- Vector Similarity $\mathrm{score\_vec}(m, c) = \frac{v_{\mathbf{source}} \cdot v_{\mathbf{target}}}{|v_{\mathbf{source}}| \cdot |v_{\mathbf{target}}|}$

- Type Compatibility $P_{\mathrm{KB}}(t) = \frac{\sum_{l \in \mathrm{map}^{-1}(t)} P_{\mathrm{NER}}(l)}{\sum_{t'} \sum_{l \in \mathrm{map}^{-1}(t')} P_{\mathrm{NER}}(l)}$

$$\mathrm{type\_compat}(m, c) = \sum_{t \in \mathrm{compatible}(t_c)} P_{\mathrm{KB}}(t)$$

Key measures to identify candidates and to perform linking

Transparent Linking

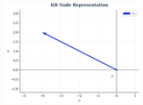
Project Overview - Entity Resolution Pipeline

Entity Recognition

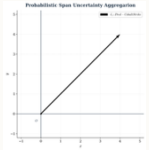
Multi Span

Multi Type

Knowledge Base Representation

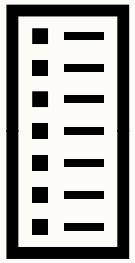


Entity Representation



Entity. | Neighbour

Candidate Generation



Informed Linking

Final Resolution

The final resolution is determined based on the *weighted* rankings.

- Vector similarity
- Lexical Score
- Type Compatibility
- Coherency

Final Resolution

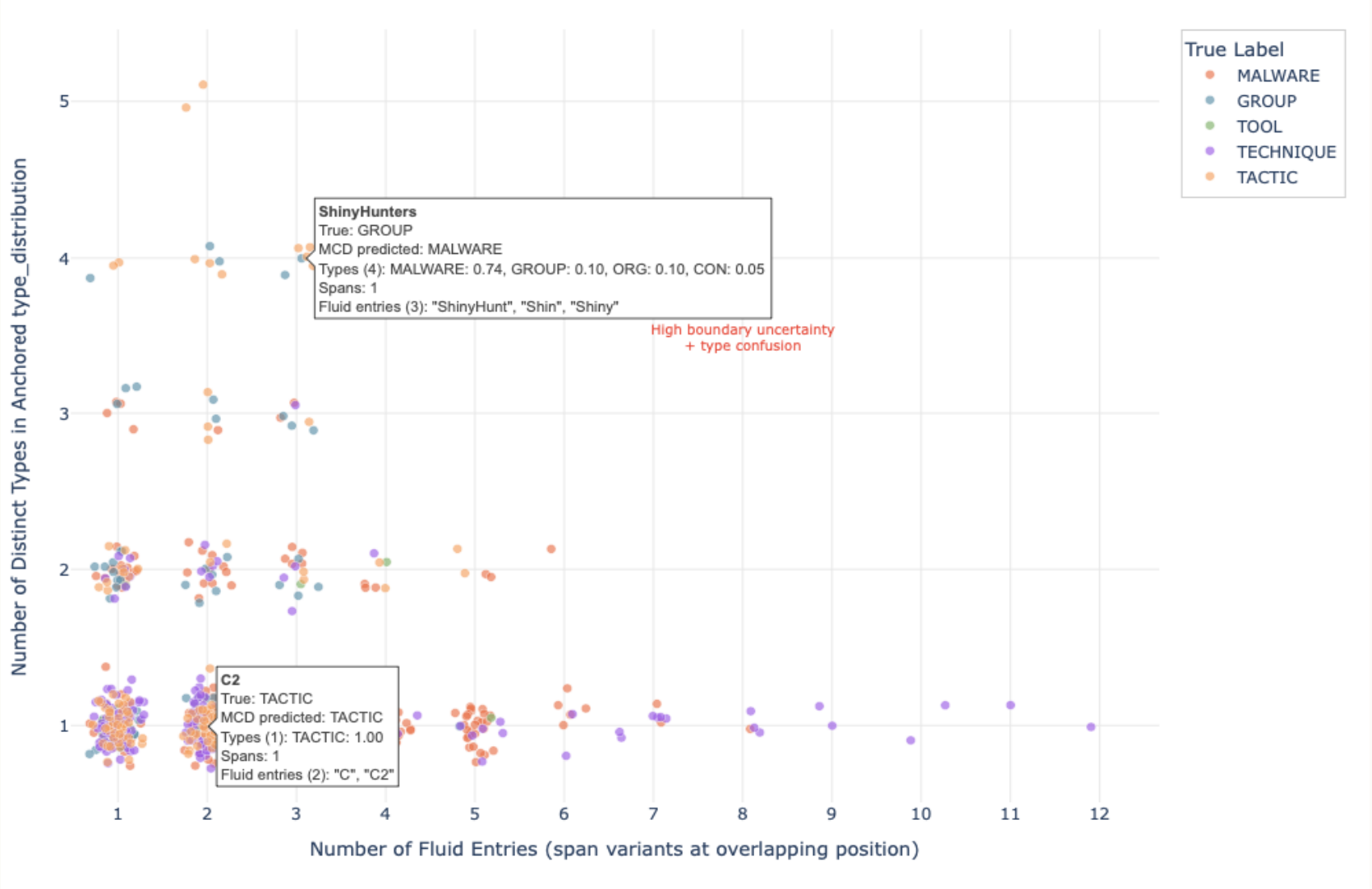
Informed Linking

Final Resolution

The aim is a better-informed resolution by

- capturing uncertainty at the beginning,
- propagating it to downstream tasks,
- resolving links transparently.

Results


True Label
MALWARE
GROUP
TOOL
TECHNIQUE
TACTIC
ShinyHunters
True: GROUP
MCD predicted: MALWARE
Types (4): MALWARE: 0.74, GROUP: 0.10, ORG: 0.10, CON: 0.05
Spans: 1
Fluid entries (3): "ShinyHunt", "Shin", "Shiny"
High boundary uncertainty
+ type confusion
C2
True: TACTIC
MCD predicted: TACTIC
Types (1): TACTIC: 1.00
Spans: 1
Fluid entries (2): "C", "C2"
Number of Distinct Types in Anchored type_distribution
Number of Fluid Entries (span variants at overlapping position)
1
2
3
4
5
6
7
8
9
10
11
12

Results

Entity Type	Gold	NER (det.)	Percentage
MALWARE	421	233	66%
GROUP	211	56	26%
TOOL	52	5	10%
TECHNIQUE	356	229	64%
TACTIC	179	125	66%

Eval Set	Correct	Incorrect	Unlinked	Precision	Recall	F1	R@1	R@5	R@10	R@30
468	199	130	139	0.605	0.425	0.499	0.397	0.434	0.442	0.453

Type	Eval	Correct	Incorrect	Unlinked	Precision	Recall	F1	R@1	R@5	R@10
MALWARE	172	154	7	11	0.957	0.895	0.925	149	154	155
GROUP	28	11	4	13	0.733	0.393	0.512	4	12	12
TOOL	2	2	0	0	1	1	1	2	2	2
TECHNIQUE	153	15	36	102	0.294	0.098	0.147	15	19	19
TACTIC	104	12	82	10	0.128	0.115	0.121	12	14	15

Methodologies

* Probabilistic Weighted Span with soft Type Filtering
* Uncertain Embeddings with pyNNDescent

Project Overview - Entity Resolution Pipeline

Entity Recognition

Knowledge Base Representation

Candidate Generation

Informed Linking

Uncertain Embeddings

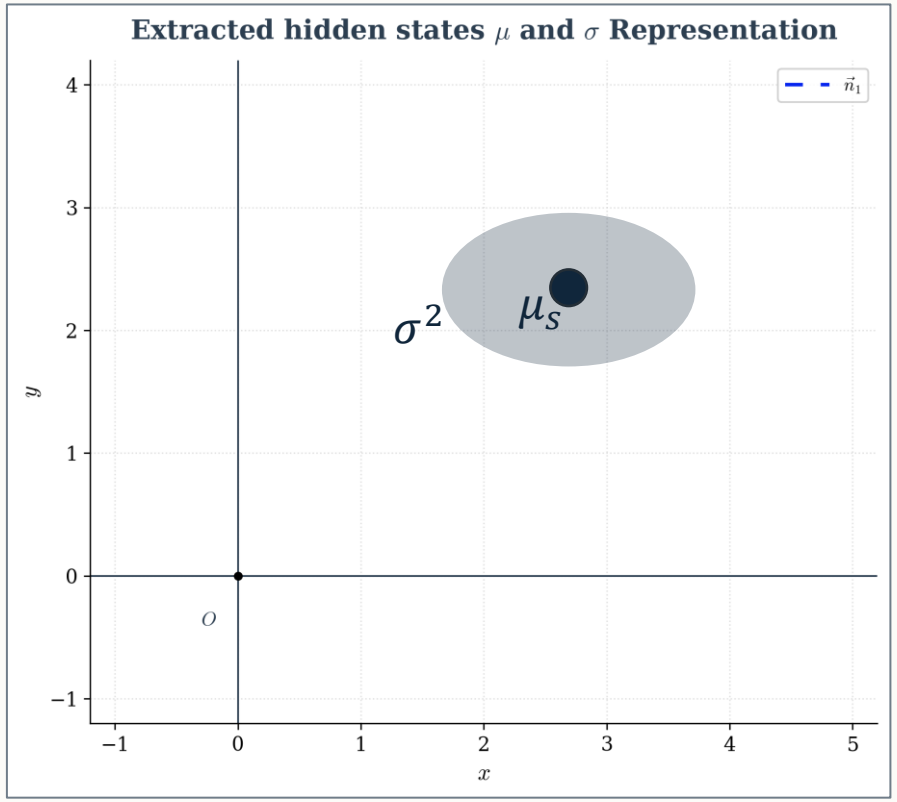
Backbone: BERT Encoder

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$s \sim \mathcal{N}\left(\mu_s, \mathrm{diag}(\sigma_s^2)\right)$$

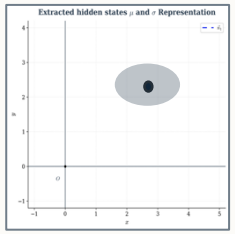
$$\mu_s \in \mathbb{R}^{768}$$

$$\sigma^2 \in \mathbb{R}^{768}$$



Project Overview - Entity Resolution Pipeline

Entity Recognition



Knowledge Base Representation

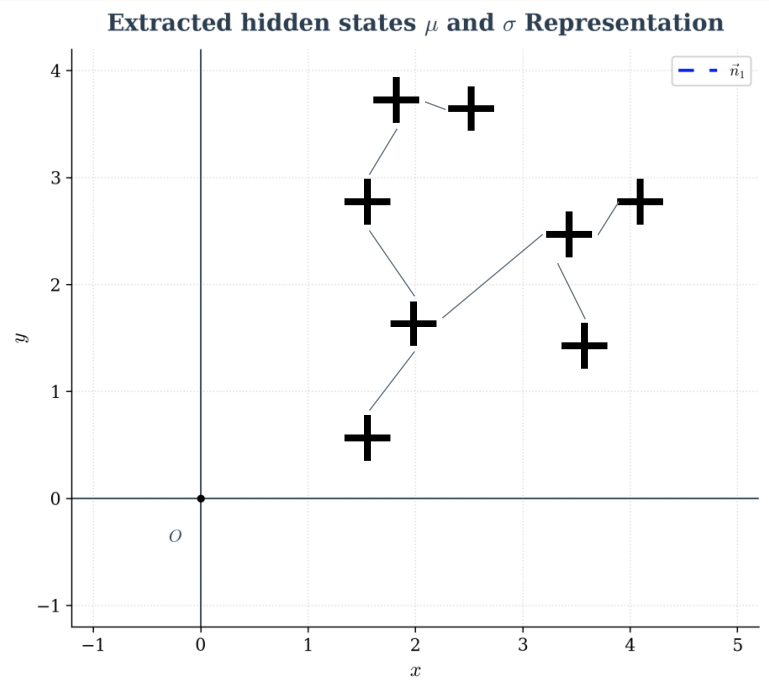
Candidate Generation

Informed Linking

KB Embeddings pyNNDescent

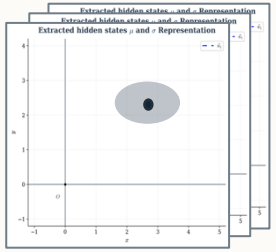
Knowledge Base Representation

Backbone: BERT Encoder

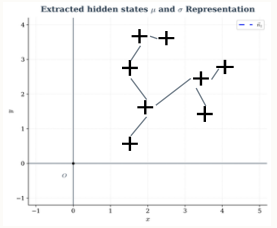


Project Overview - Entity Resolution Pipeline

Entity Recognition



Knowledge Base Representation

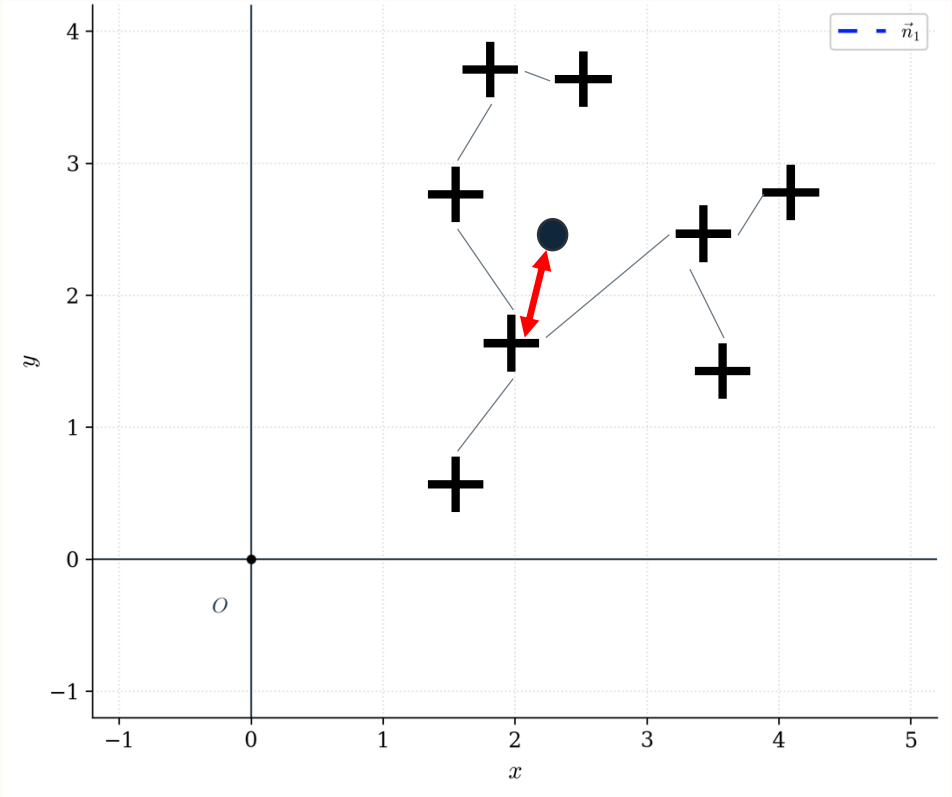


Candidate Generation

Informed Linking

Uncertain Embedding pyNNDescent

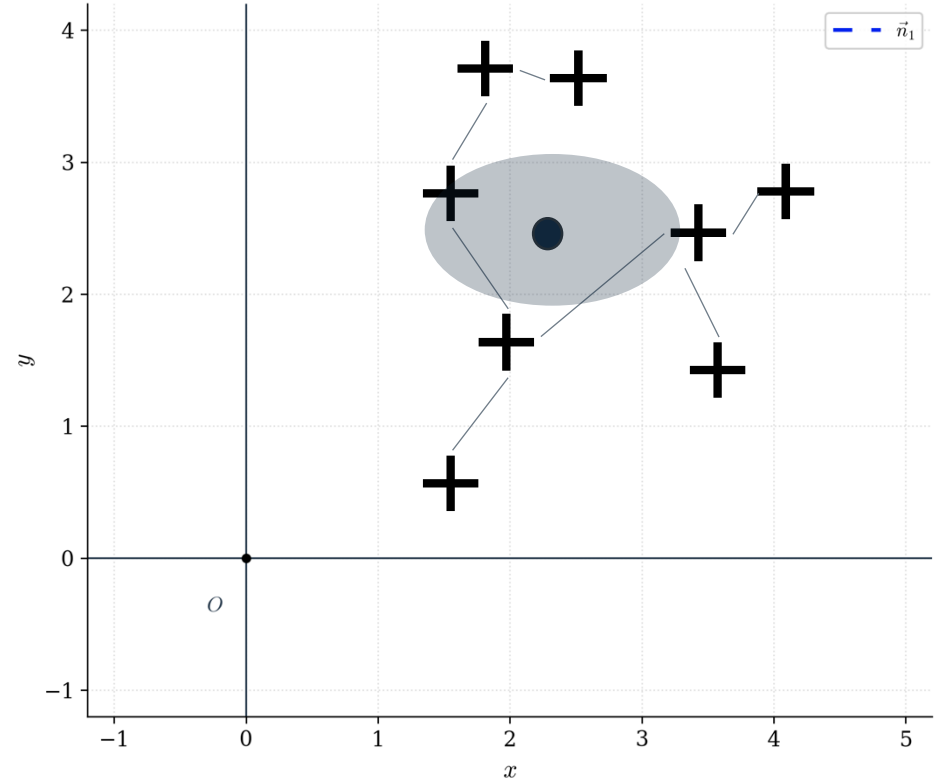
Candidate Generation



Node embedding

Source embedding

Uncertain Embedding pyNNDescent



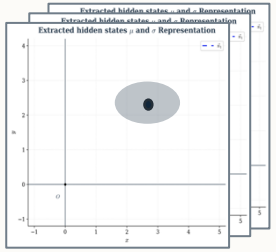
μ_t Node embedding

μ_s Source embedding

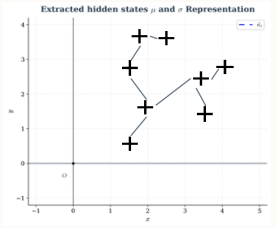
$$d(s,t) = \sum_{i=1}^{D} \frac{\left(\mu_{s,i} - \mu_{t,i}\right)^2}{\sigma_{s,i}^2 + \epsilon}$$

Project Overview - Entity Resolution Pipeline

Entity Recognition



Knowledge Base Representation

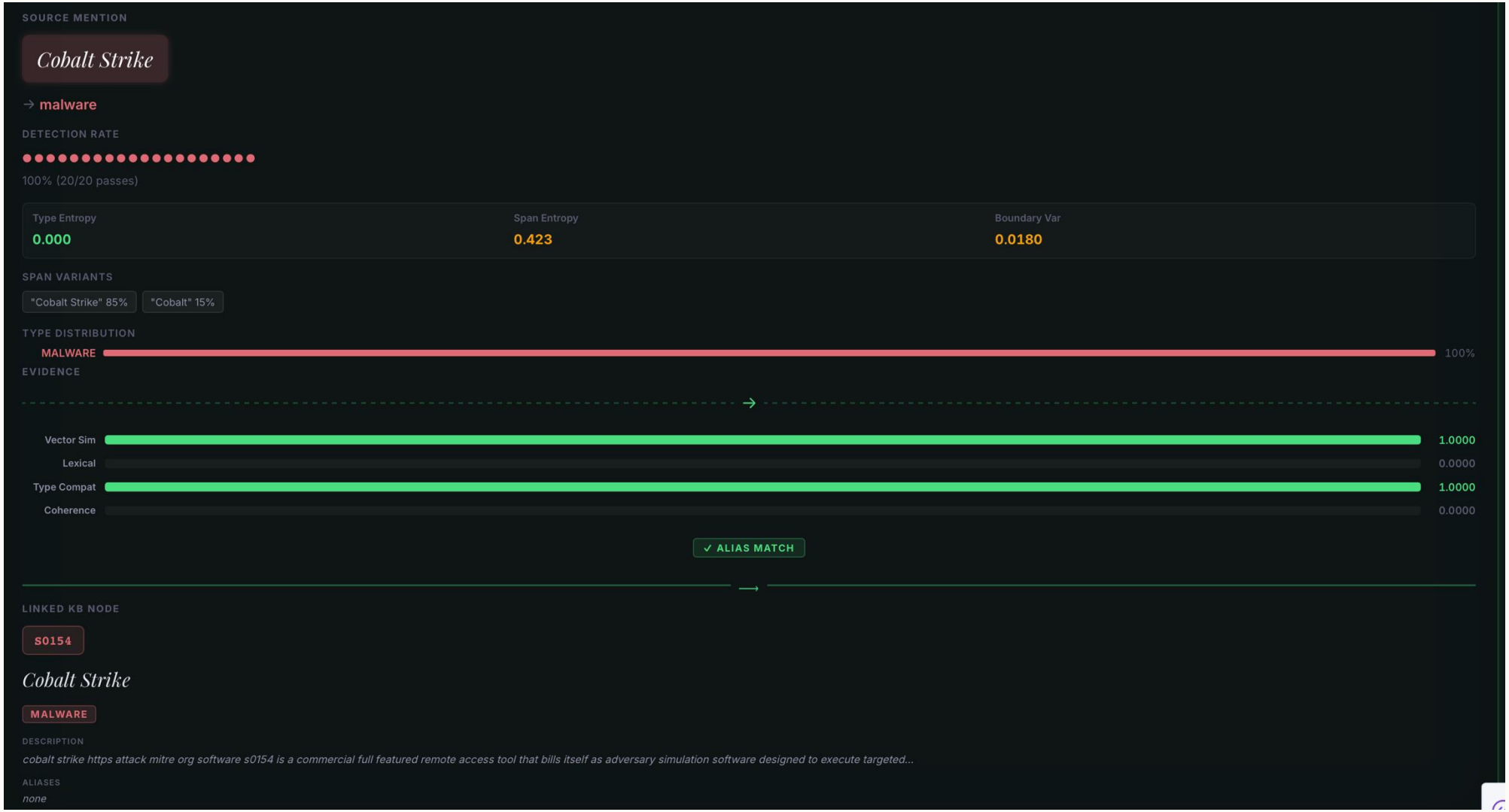


Candidate Generation

Informed Linking

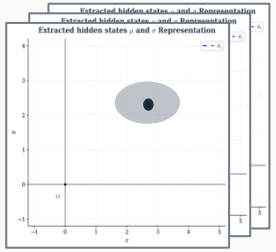
Resolution

Informed Linking

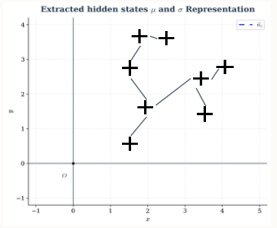


Entity Resolution Pipeline

Entity Recognition



Knowledge Base Representation



Candidate Generation

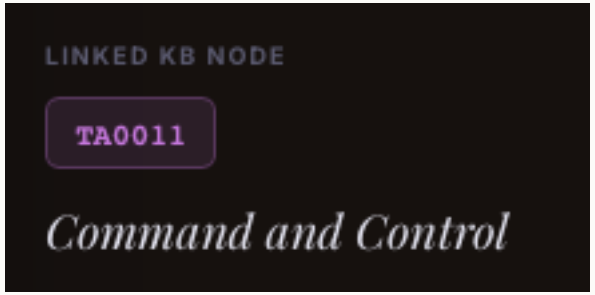
Informed Linking

Results


SOURCE MENTION
Visual Basic Script
→ technique


LINKED KB NODE
T1059
Command and Scripting Interpreter
TECHNIQUE

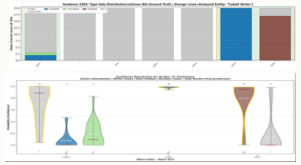

SOURCE MENTION
CnC
→ tactic

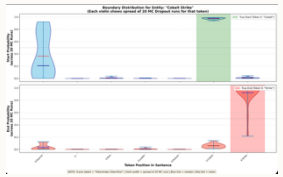

LINKED KB NODE
TA0011
Command and Control

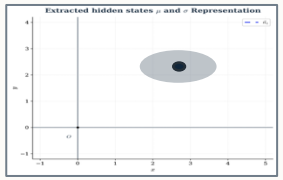
Future Work

- Uncertainty understanding
- Aggregation

Uncertainty Understanding Decomposition





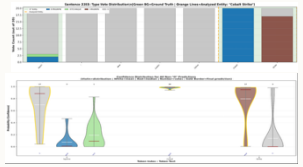


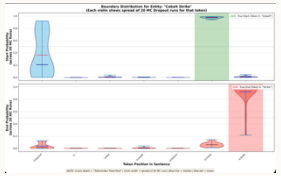
What remains for uncertainty capture?

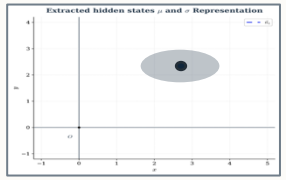
So far we mainly capture model instability.

The next step is to understand *what kind* of uncertainty we are dealing with and why it arises.

Uncertainty Understanding Aggregation



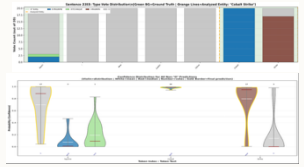


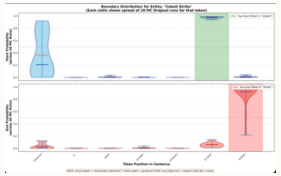


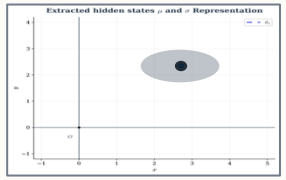
Aggregation is not neutral — it actively shapes downstream decisions.

- Should all uncertainty dimensions be treated equally?
- How should uncertainty be aggregated ?
- When does aggregation hide useful signals?

Decision-Maker Engagement





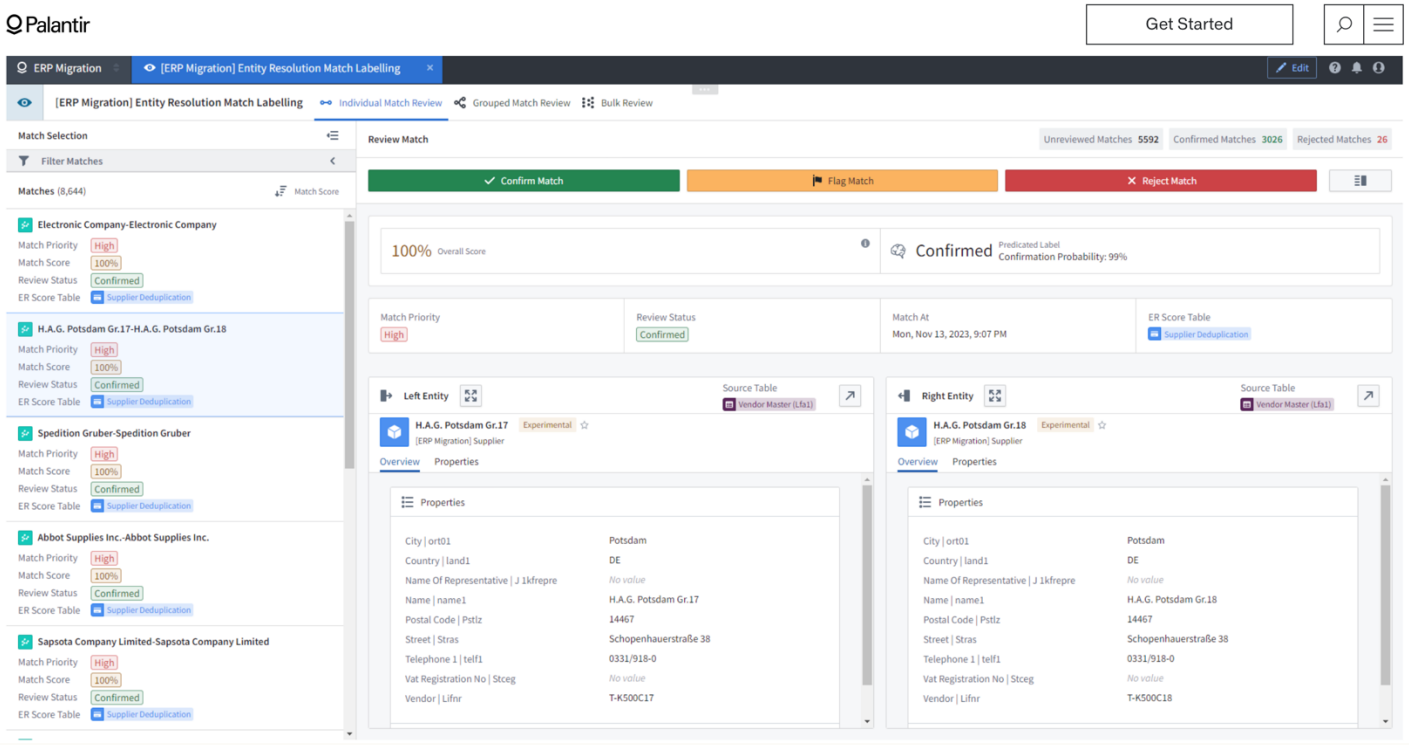


Co-Design / Decision-Maker Engagement

How do analysts interpret different uncertainty signals?

When does uncertainty help vs overwhelm?

Which representations support trust and action?



Q & A

direnc.pekaslan1@nottingham.ac.uk