





NIKOLAOS I. SPANOUDAKIS¹, ANTONIS C. KAKAS² AND ADAMOS KOUMI²

¹SCHOOL OF PRODUCTION ENGINEERING AND MANAGEMENT, TECHNICAL UNIVERSITY OF CRETE

²DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF CYPRUS

Outline

- Introduction Contributions
- ☐ Building explanations from Gorgias preference-based argumentation framework results
- Application examples
- Conclusion and Future Work

Contributions

- We show how the returned results of the dialectical argumentation reasoning within the Gorgias framework can be exploited to provide human-readable explanations that are
 - Attributive
 - Contrastive
 - Actionable
- ☐ These results, can be manipulated by applications to produce case-based human readable explanations.

A gorgias theory example (code)

```
rule(r1(X), buy(X), []):- need(X).
rule(r2(X), neg(buy(X)), [neg(urgentNeed(X))]).
rule(pr1(X), prefer(r2(X), r1(X)), [lowOnFunds]).
rule(pr2(X), prefer(r1(X), r2(X)), []).
rule(c1(X), prefer(pr1(X), pr2(X)), []).
abducible(urgentNeed(X), []).
abducible(neg(urgentNeed(X)), []).
```

A gorgias query example

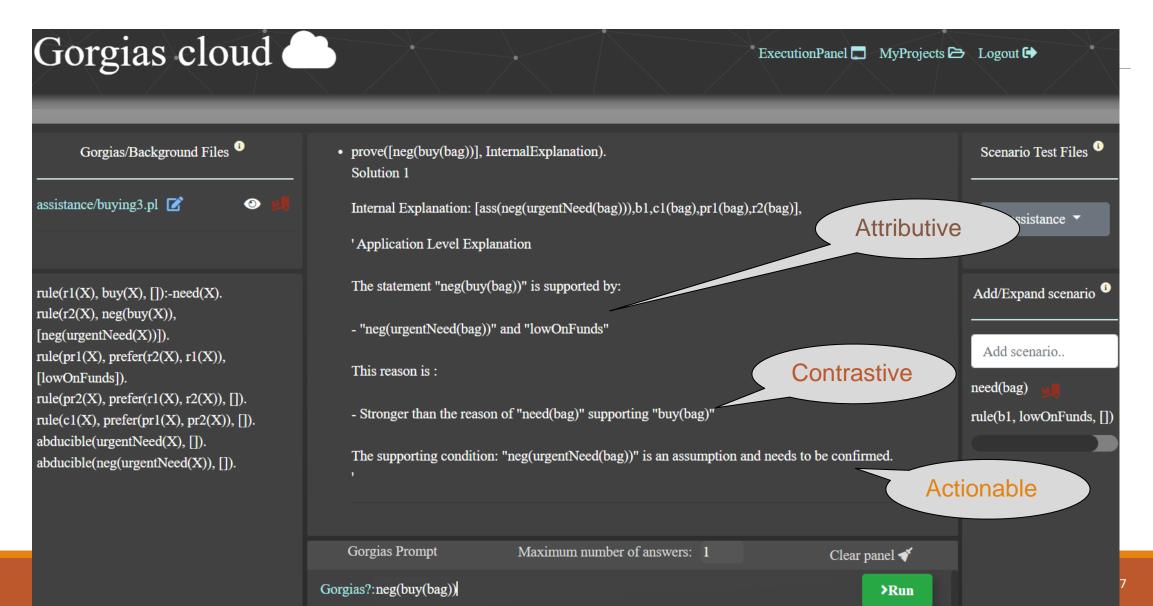
- □ Suppose that we know that the agent needs a bag and it believes it is low on funds
 - need(bag)
 - lowOnFunds -> rule(b1, lowOnFunds, [])
- Let's try the query to not buy the bag
 - neg(buy(bag))
- ☐ It is valid, the argument is
 - ass(neg(urgentNeed(bag))), c1(bag), b1, pr1(bag), r2(bag)

Actionable

Explanation generator

- Internal explanation:
- □ b1, r2(bag), pr1(bag), c1(bag), ass(neg(urgentNeed(bag)))
 - rule(r1(X), buy(X), []):- need(X).
 - rule(r2(X), neg(buy(X)), [neg(urgentNeed(X))]).
 - rule(pr1(X), prefer(r2(X), r1(X)), [lowOnFunds]).
 - rule(pr2(X), prefer(r1(X), r2(X)), []).
 - rule(c1(X), prefer(pr1(X), pr2(X)), []).
 - abducible(urgentNeed(X), []).
 - abducible(neg(urgentNeed(X)), []).

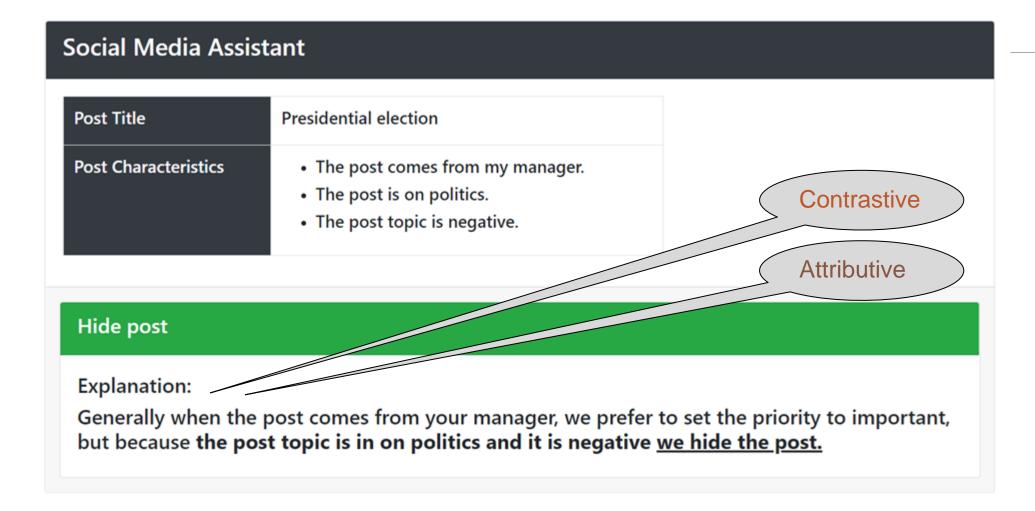
Gorgias Cloud Application Level Explanations



A Social Media Application Example

- ☐ The agent browses social media content
- Sets each item's priority to:
 - Important
 - Normal/default
 - Hide
- Policy:
 - ... Posts that come from the user's manager are important regardless of whether they are positive or negative. ... Hide politics posts from the user's manager when negative. ...

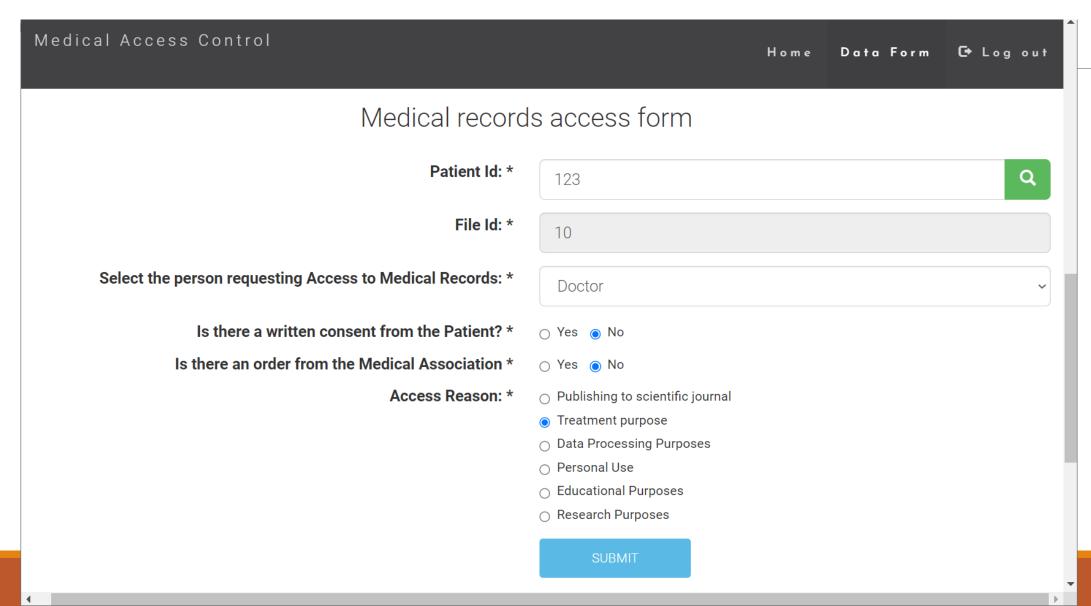
A Social Media Application Example



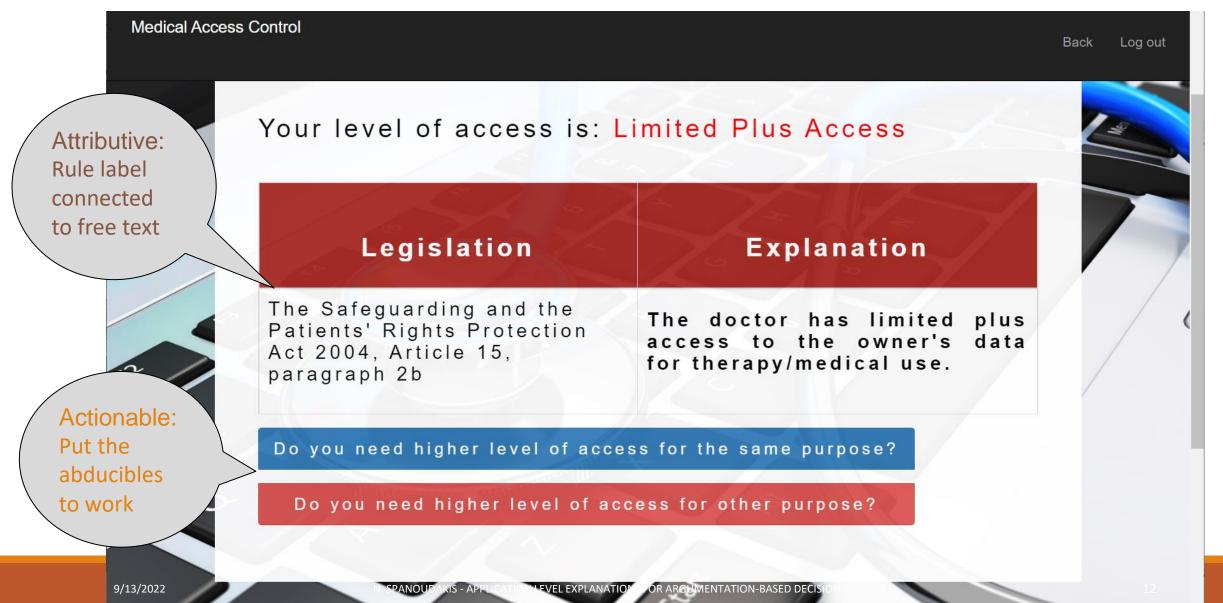
Application Example 1: Medica

- ☐ This system aims to aid the decisions of administrative personnel in the health domain
- They need to decide what information can be disclosed to a person asking for it
- ☐ EU legislation defines the access rights

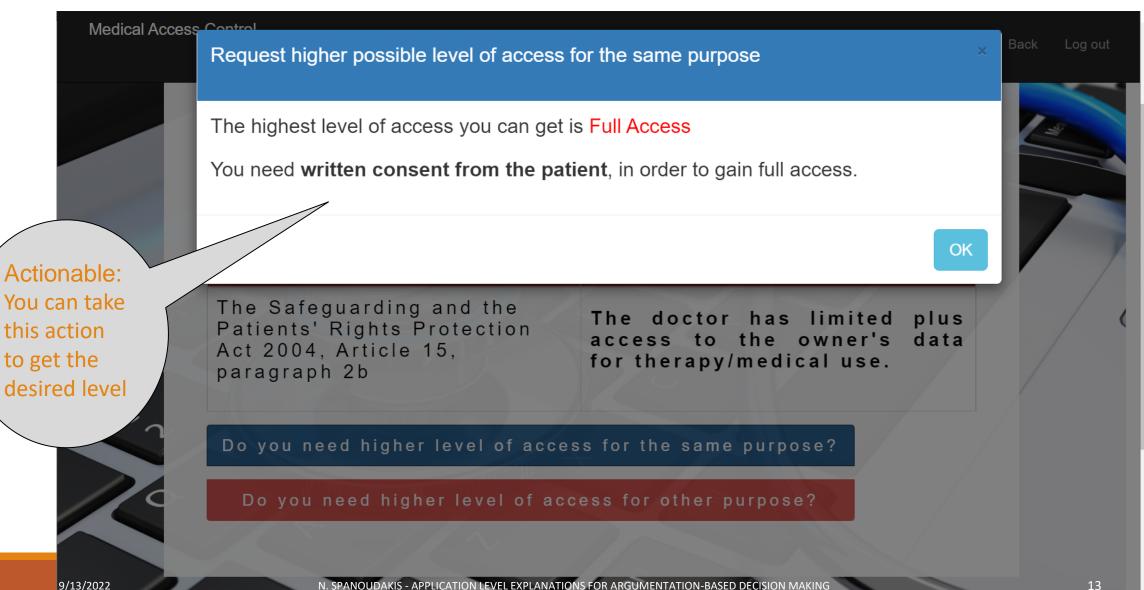
Application Example 1: Medica



Application Example 1: Medica



Application Example 1: Medica\



Application Example 2: GAID

- ☐ GAID: Gynecological AI Diagnostic Assistant
- ☐ This system aims to aid a medical doctor in disease diagnosis
- ☐ The system gets the tests and symptoms of a patient
- Doctor's knowledge determines the outcome

Application Example 2: GAID



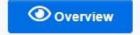


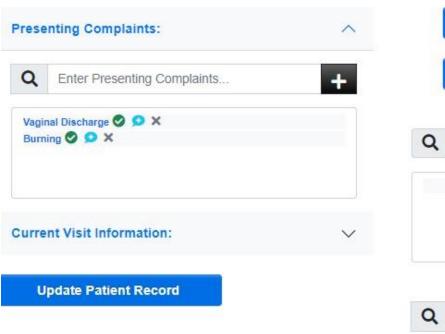


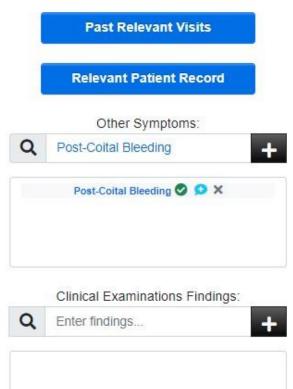
Gynaecological Artificial Intelligence Diagnostics - Cognitive Assistant

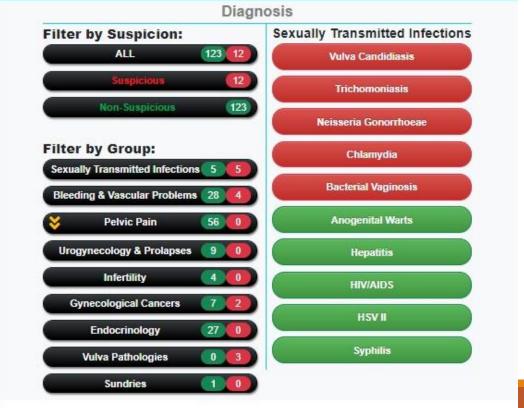
Current Patient: Sylvia Tina Jackson Heath (ID: 5544332211)

Current Date: Aug 11 2021, Wed









Application Example 2: GAID

Diagnosis:

Anogenital Warts

Attributive

EXPLANATION

RELATED INFO

ICD-10

Under the information Vaginal Burning it is recommended that you investigate Anogenital Warts.

Contrastive

This decision is supported by: Inter-Menstrual Bleeding.

The following further information strengthens this decision: Dysuria.

But each of the following information:

- Image: No indication of small cauliflower-shaped lumps
- History: Vaccinated with HPV before first intercourse

indicates, that this disease may not be possible and could be excluded.

Actionable:
Possible issues to investigate in order to exclude other possible diseases

Conclusions

- ☐ We have delved into explainable AI-base decisions that are
 - Attributive
 - Contrastive
 - Actionable
- Still more work is needed
 - NLP: generate predicates and arguments from human generated text
 - Generate arguments explanations (Attributive, Contrastive or Actionable) in free text

Thank you, questions?

Nikos Spanoudakis

Researcher - Teaching Staff
Applied Mathematics and Computers Laboratory,
School of Production Engineering and Management,
Technical University of Crete,

Email: nikos@amcl.tuc.gr

https://users.isc.tuc.gr/~nispanoudakis



The Gorgias Cloud System will be a demo at COMMA 22 (Wednesday)

Open for academic use: http://gorgiasb.tuc.gr/GorgiasCloud.html