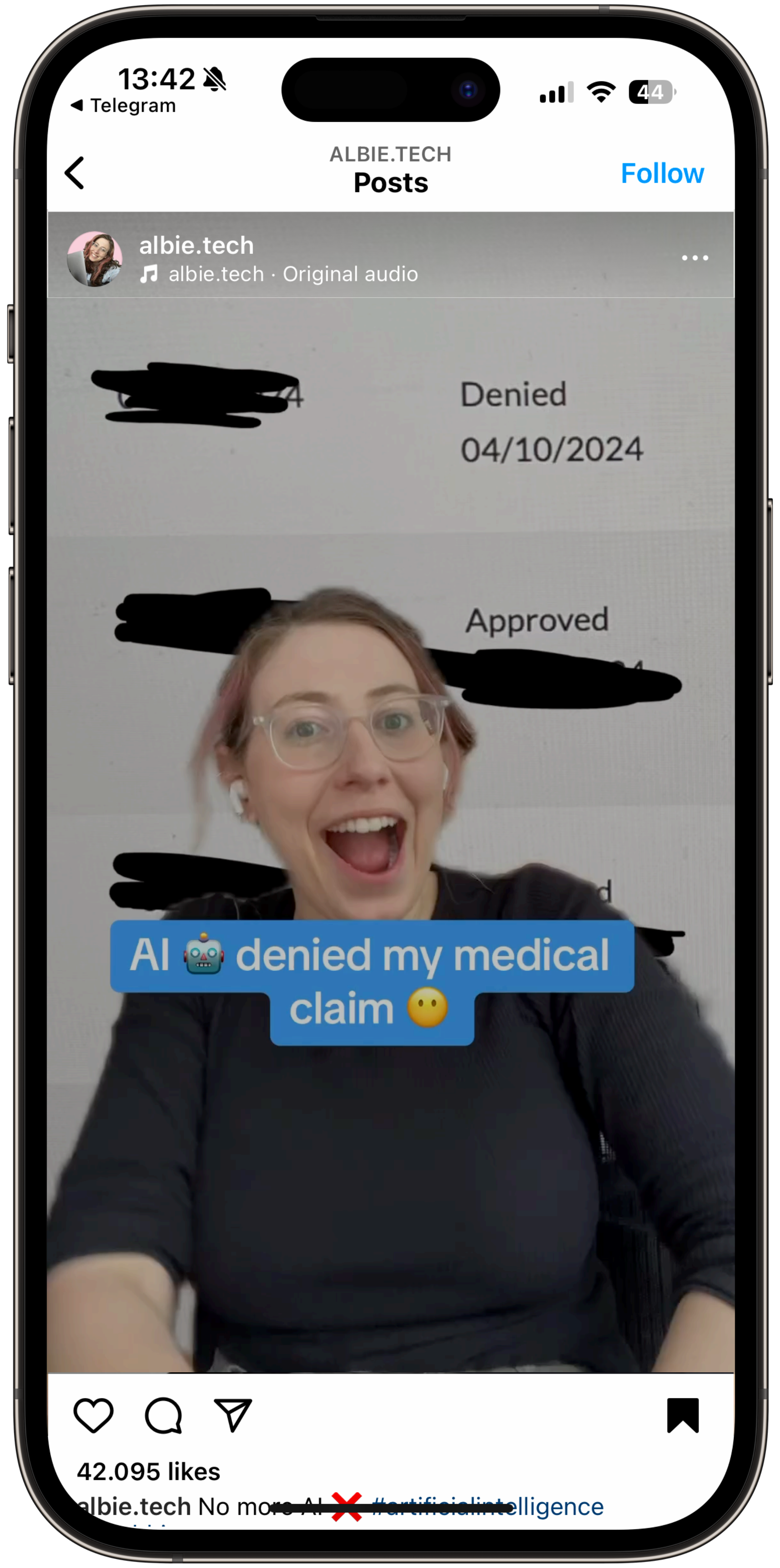
Explainable AI

Marcel Lahaye: CTHCI '24



[redacted] Denied 04/10/2024

[redacted] Approved [redacted]

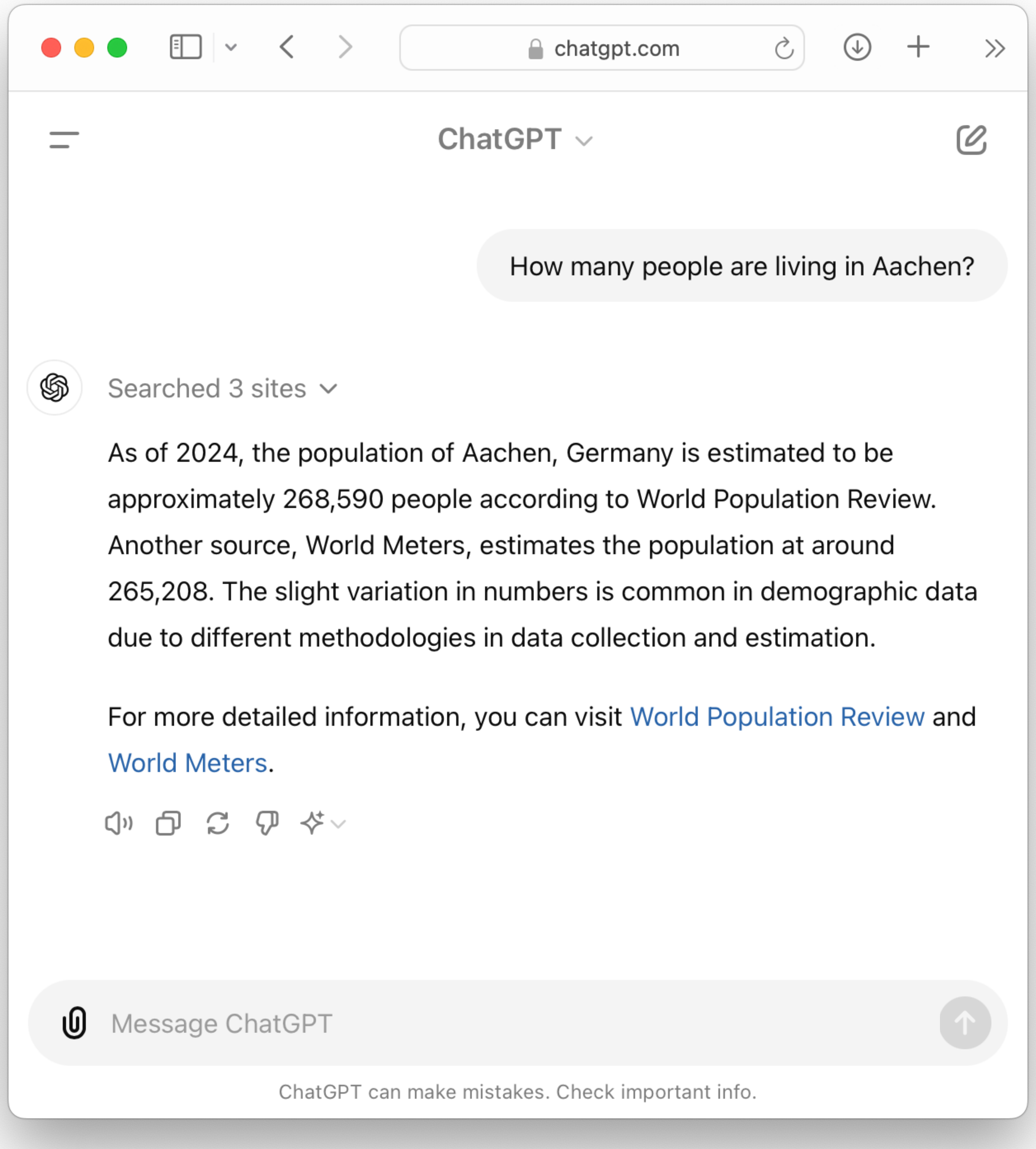
[redacted]

AI 🤖 denied my medical claim 😶

42.095 likes

albie.tech No mo~~re AI ❌ #artificialint~~elligence

@albie.tech,
“No more AI ❌ #artificialintelligence #healthinsurance”,
Instagram, April 13th 2024,
https://www.instagram.com/reel/C5tNFRTuzrX

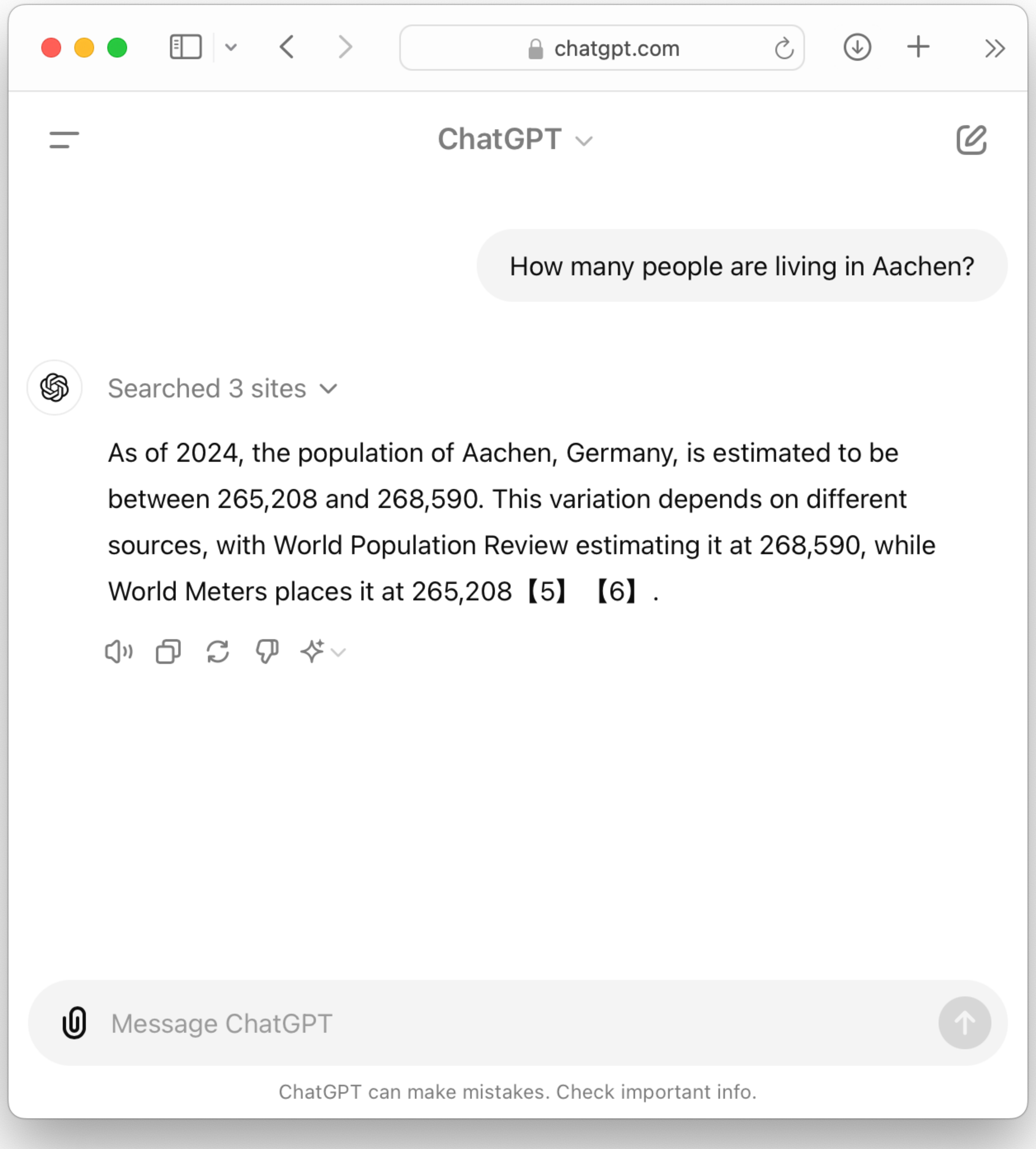


How many people are living in Aachen?

As of 2024, the population of Aachen, Germany is estimated to be approximately 268,590 people according to World Population Review. Another source, World Meters, estimates the population at around 265,208. The slight variation in numbers is common in demographic data due to different methodologies in data collection and estimation.

For more detailed information, you can visit World Population Review and World Meters.

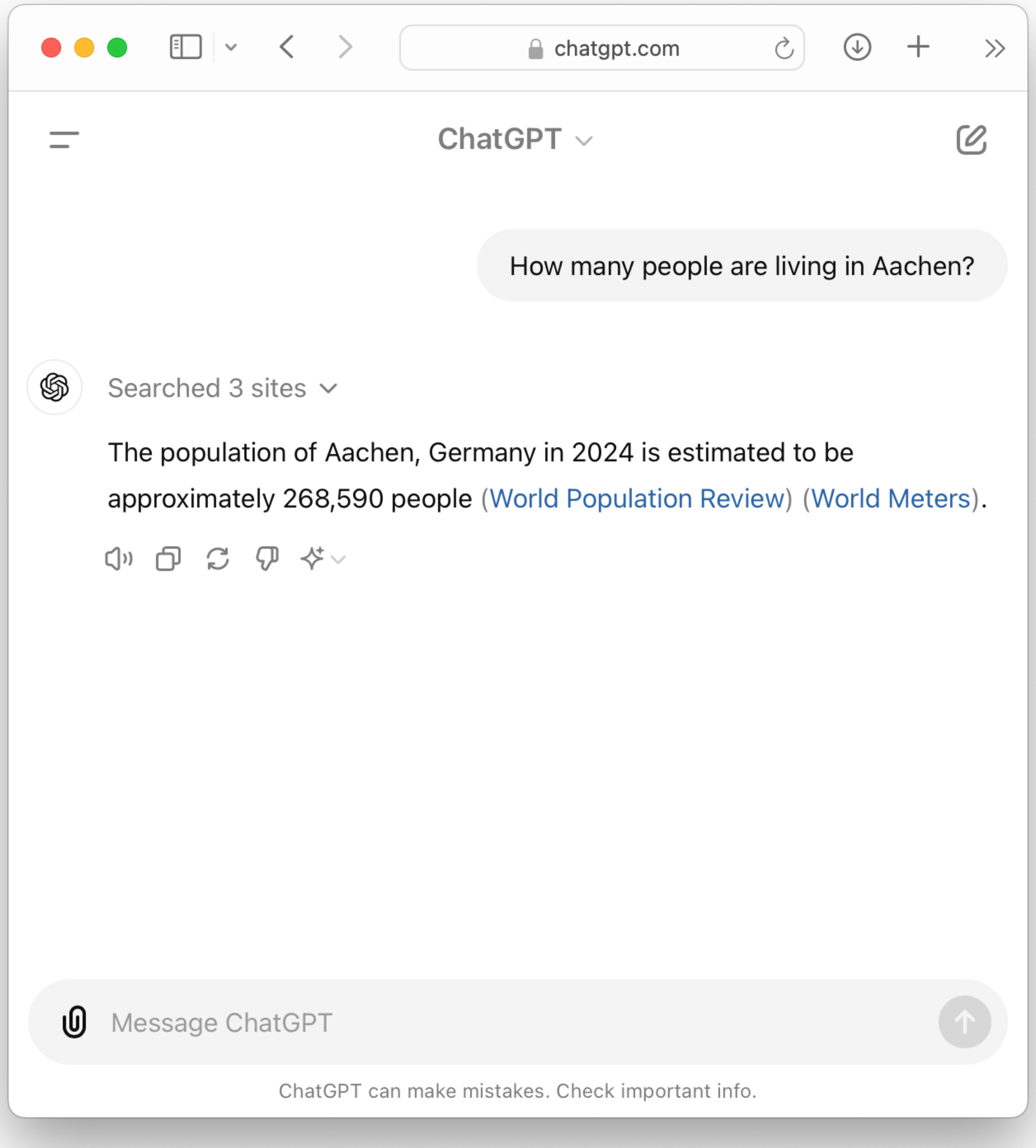
ChatGPT can make mistakes. Check important info.



How many people are living in Aachen?

As of 2024, the population of Aachen, Germany, is estimated to be between 265,208 and 268,590. This variation depends on different sources, with World Population Review estimating it at 268,590, while World Meters places it at 265,208 【5】 【6】 .

ChatGPT can make mistakes. Check important info.



How many people are living in Aachen?

The population of Aachen, Germany in 2024 is estimated to be approximately 268,590 people (World Population Review) (World Meters).

ChatGPT can make mistakes. Check important info.

OpenAI, ChatGPT, http://chatgpt.com, accessed June 21st 2024

> “Explainable AI (XAI), a research area that aims to provide human-understandable justifcations for the system’s behavior.”

- Ehsan et al., The Who in XAI: How AI Background Shapes Perceptions of AI Explanations, CHI ’24

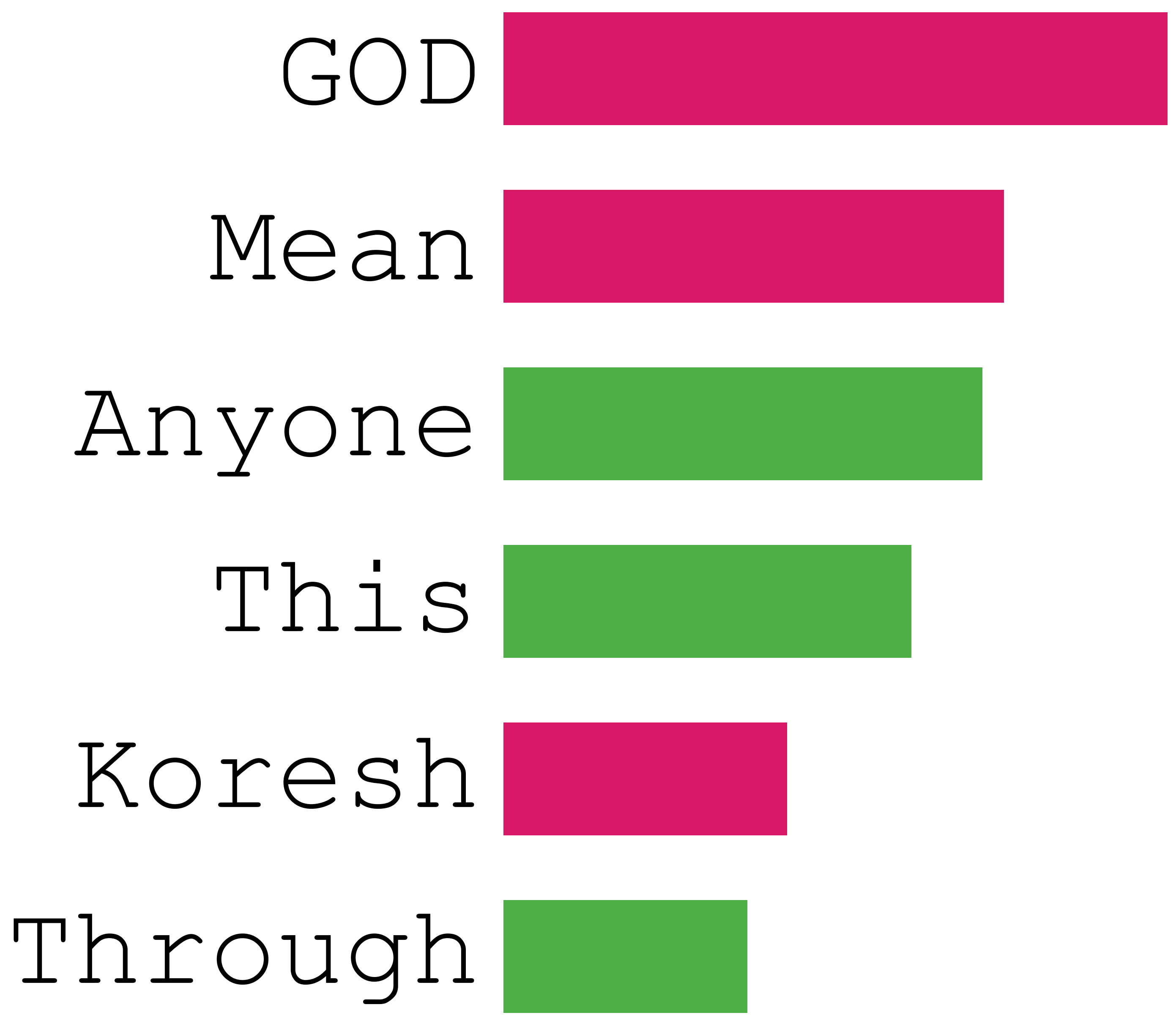
Local Interpretable Model-agnostic Explanations (LIME)

```
From: pauld@verdix.com (Paul Durbin)
Subject: Re: DAVID CORESH IS! GOD!
Nntp-Posting-Host: sarge.hq.verdix.com
Organization: Verdix Corp
Lines: 8
```

Ribeiro et al., "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD '16

Local Interpretable Model-agnostic Explanations (LIME)

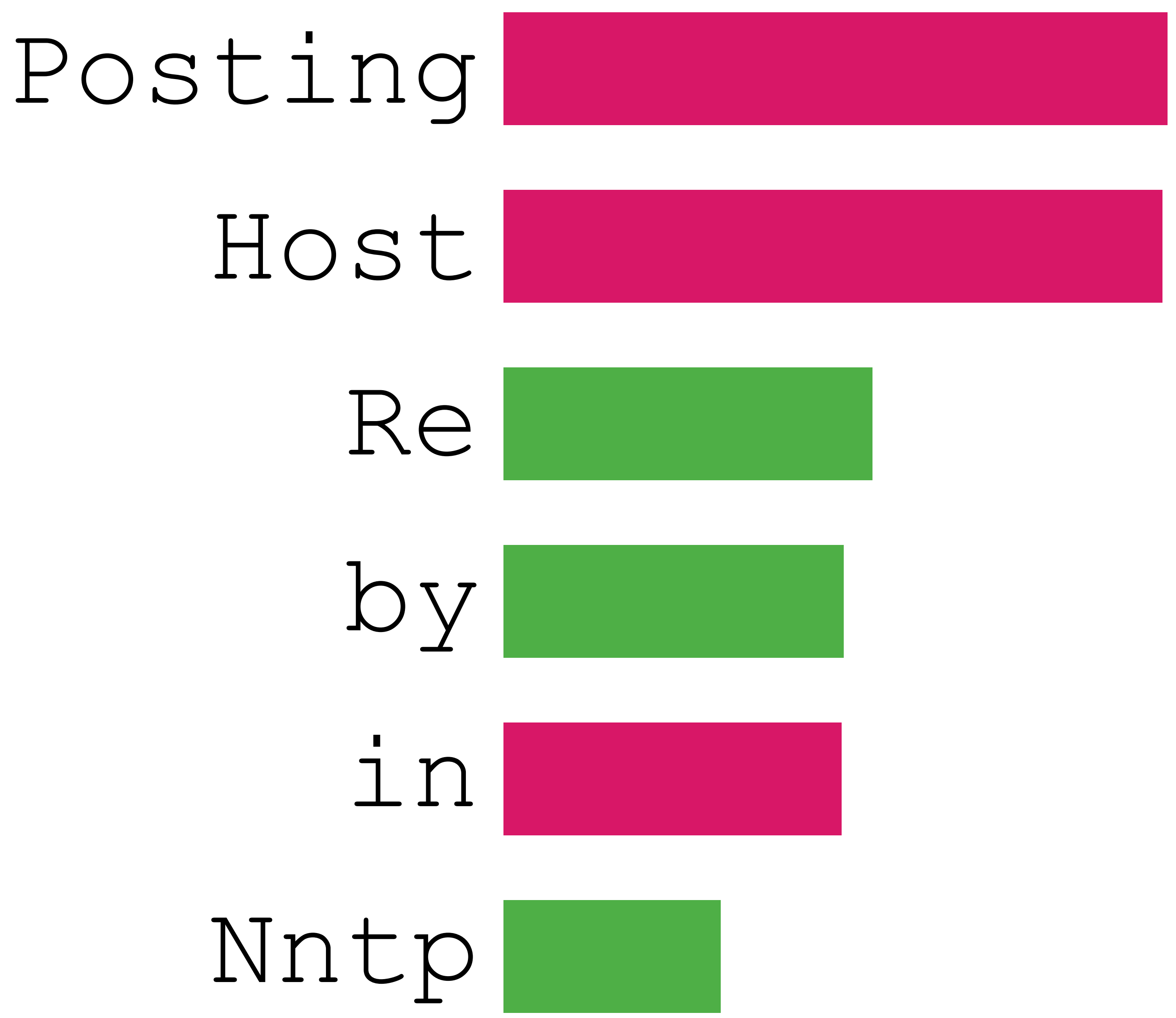
```
From: pauld@verdix.com (Paul Durbin)
Subject: Re: DAVID CORESH IS! GOD!
Nntp-Posting-Host: sarge.hq.verdix.com
Organization: Verdix Corp
Lines: 8
```



Ribeiro et al., "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD ’16

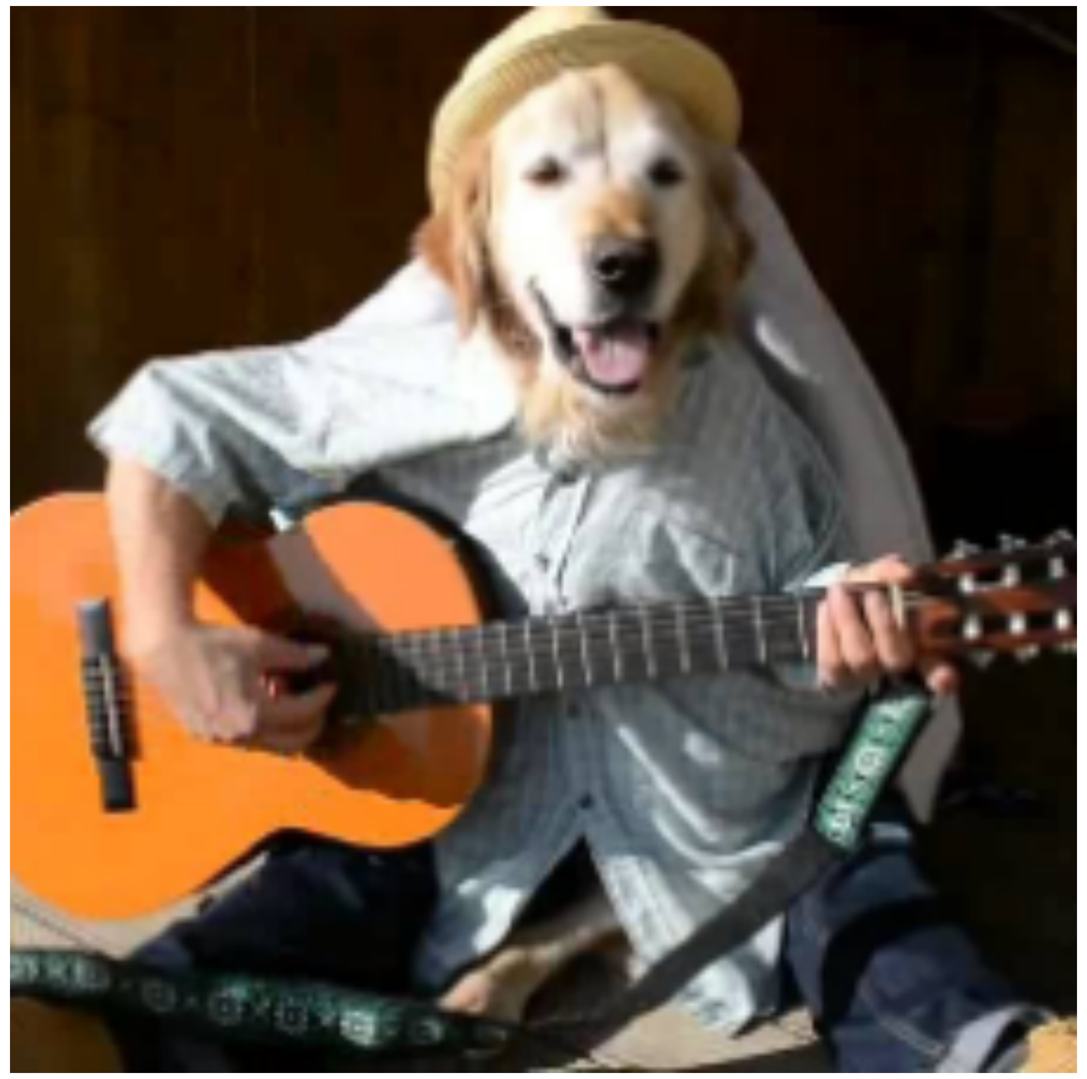
Local Interpretable Model-agnostic Explanations (LIME)

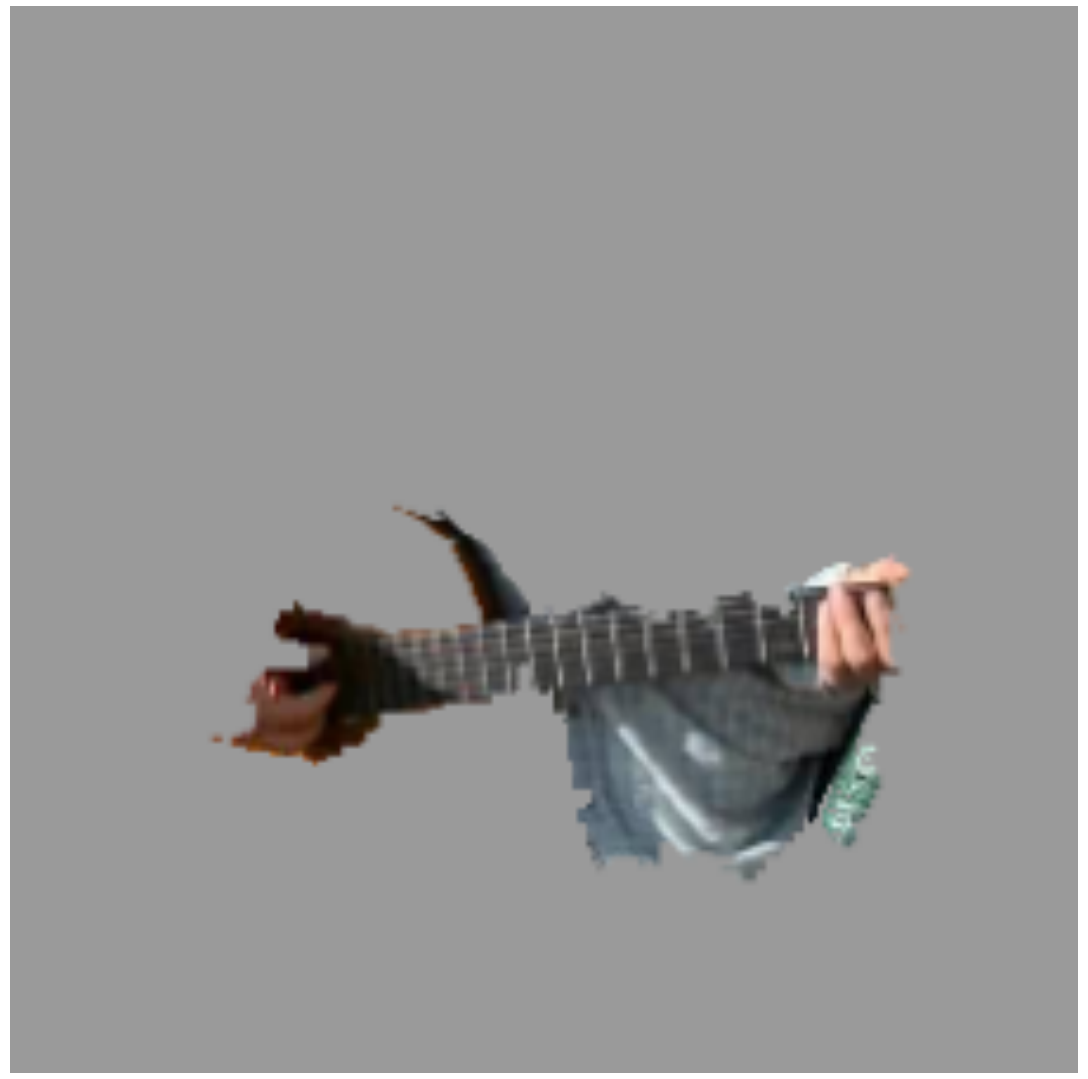
From: pauld@verdix.com (Paul Durbin)
Subject: **Re**: DAVID CORESH IS! GOD!
Nntp-Posting-Host: sarge.hq.verdix.com
Organization: Verdix Corp
Lines: 8



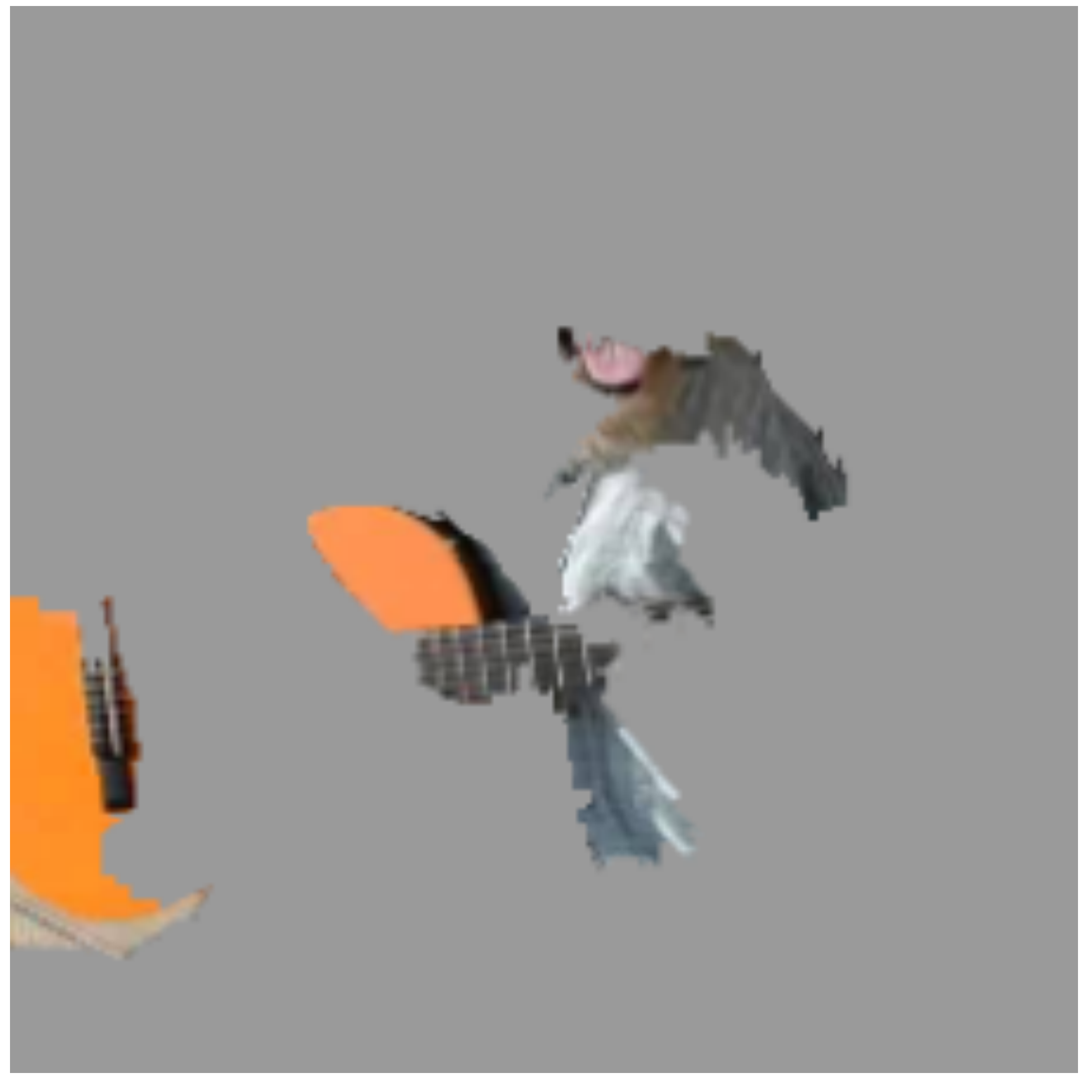
Ribeiro et al., "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD ’16

Local Interpretable Model-agnostic Explanations (LIME)

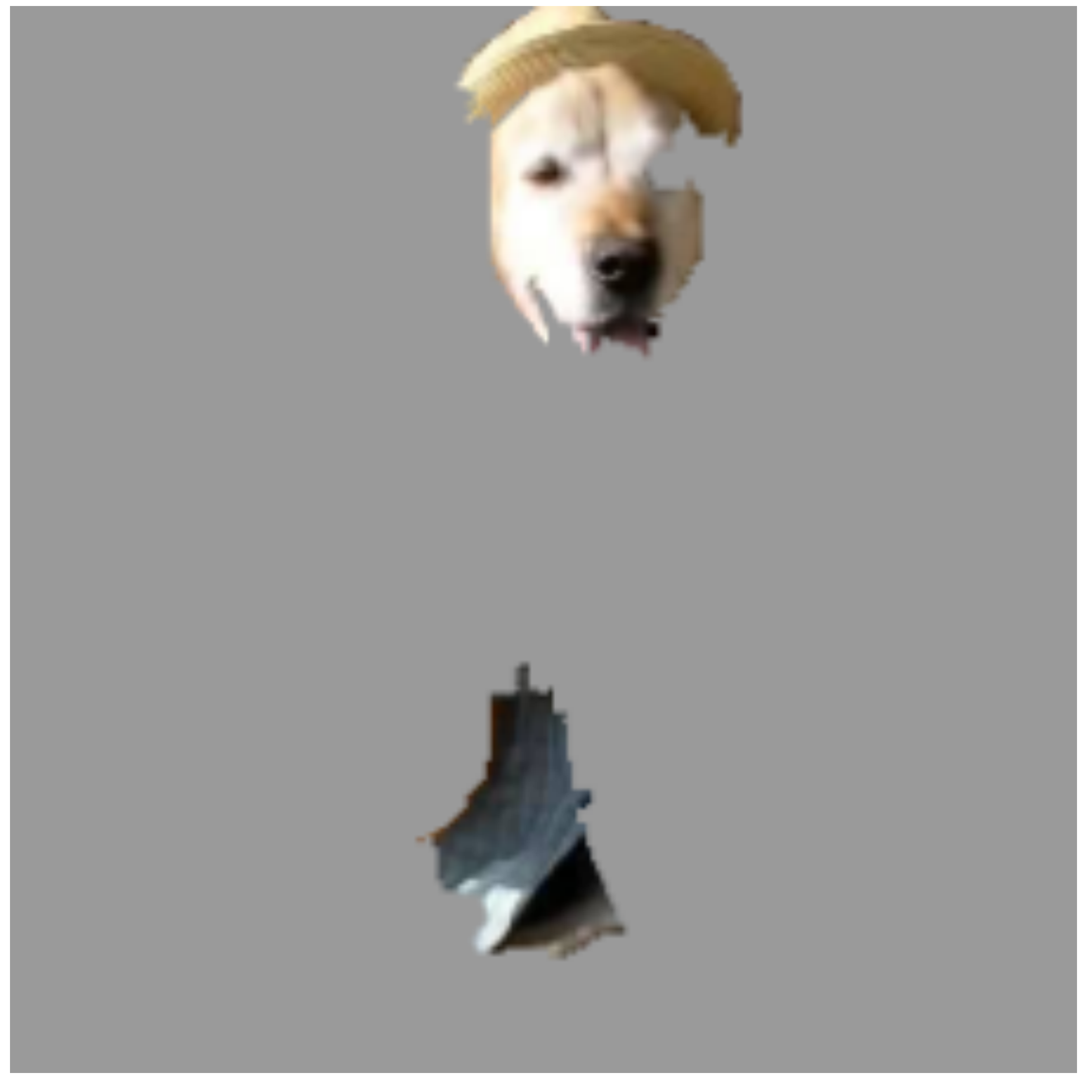




Electric Guitar



Acoustic Guitar

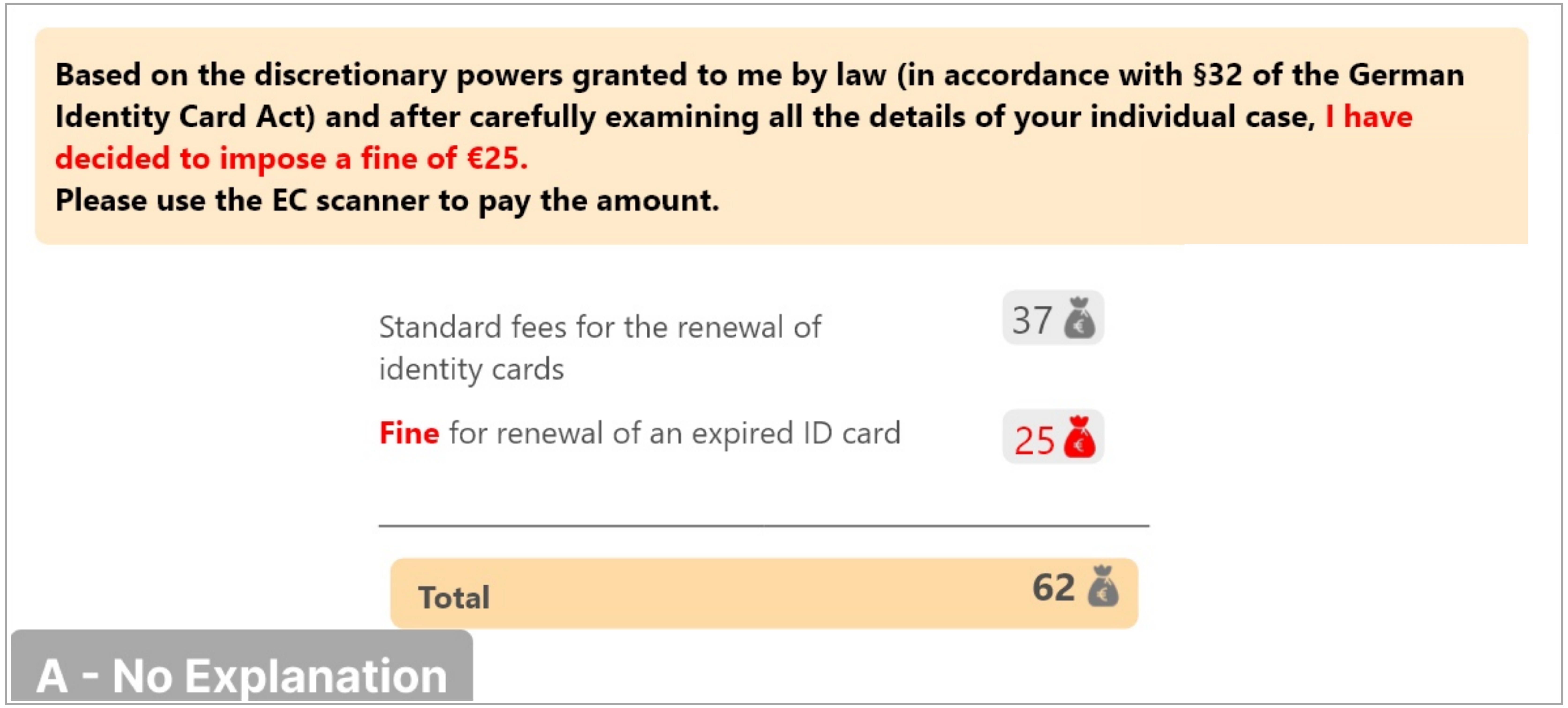


Labrador

Ribeiro et al., "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD ’16

Explanation Level

CHI '24



Based on the discretionary powers granted to me by law (in accordance with §32 of the German Identity Card Act) and after carefully examining all the details of your individual case, I have decided to impose a fine of €25.
Please use the EC scanner to pay the amount.

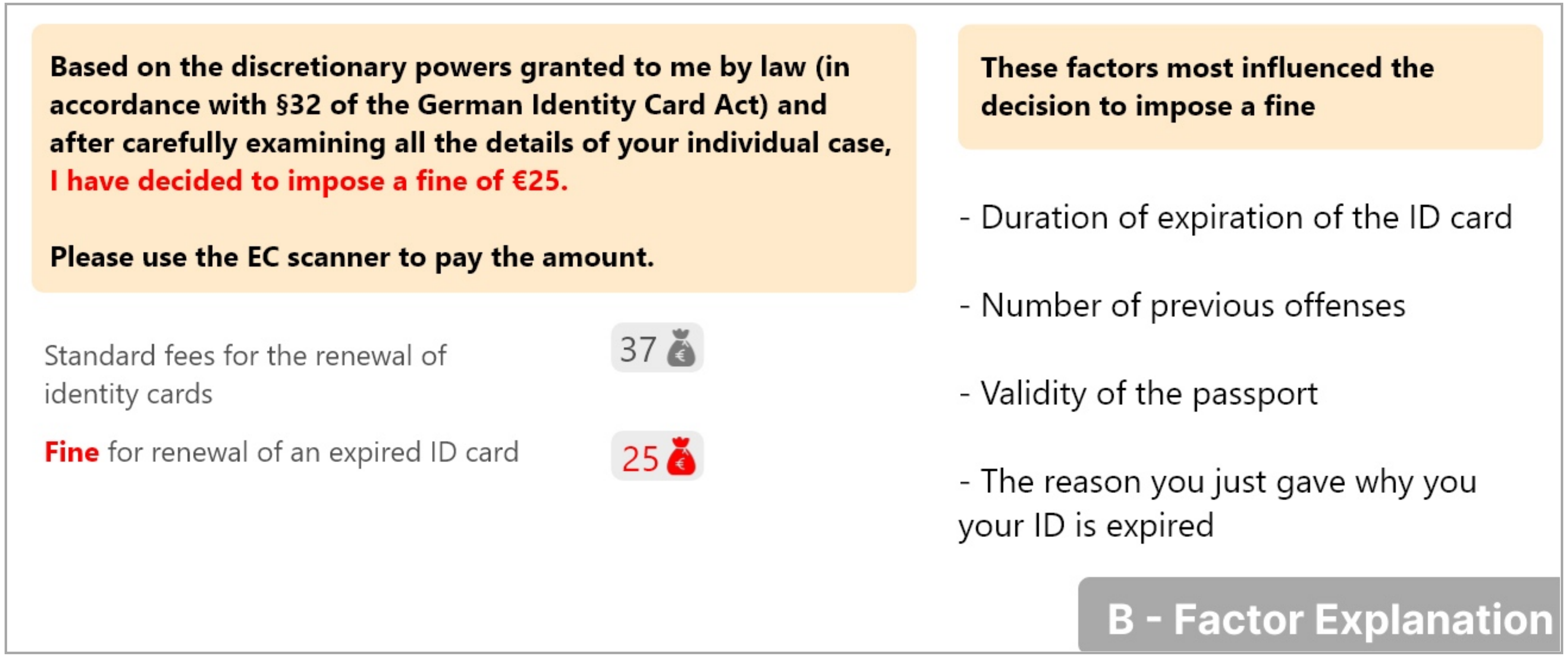
Standard fees for the renewal of identity cards	37 💰
Fine for renewal of an expired ID card	25 💰
Total	**62** 💰

A - No Explanation

Aljuneidi et al., Why the Fine, AI?
The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations, CHI '24

Explanation Level

CHI ’24



Based on the discretionary powers granted to me by law (in accordance with §32 of the German Identity Card Act) and after carefully examining all the details of your individual case, I have decided to impose a fine of €25.

Please use the EC scanner to pay the amount.

Standard fees for the renewal of identity cards	37
Fine for renewal of an expired ID card	25

These factors most influenced the decision to impose a fine

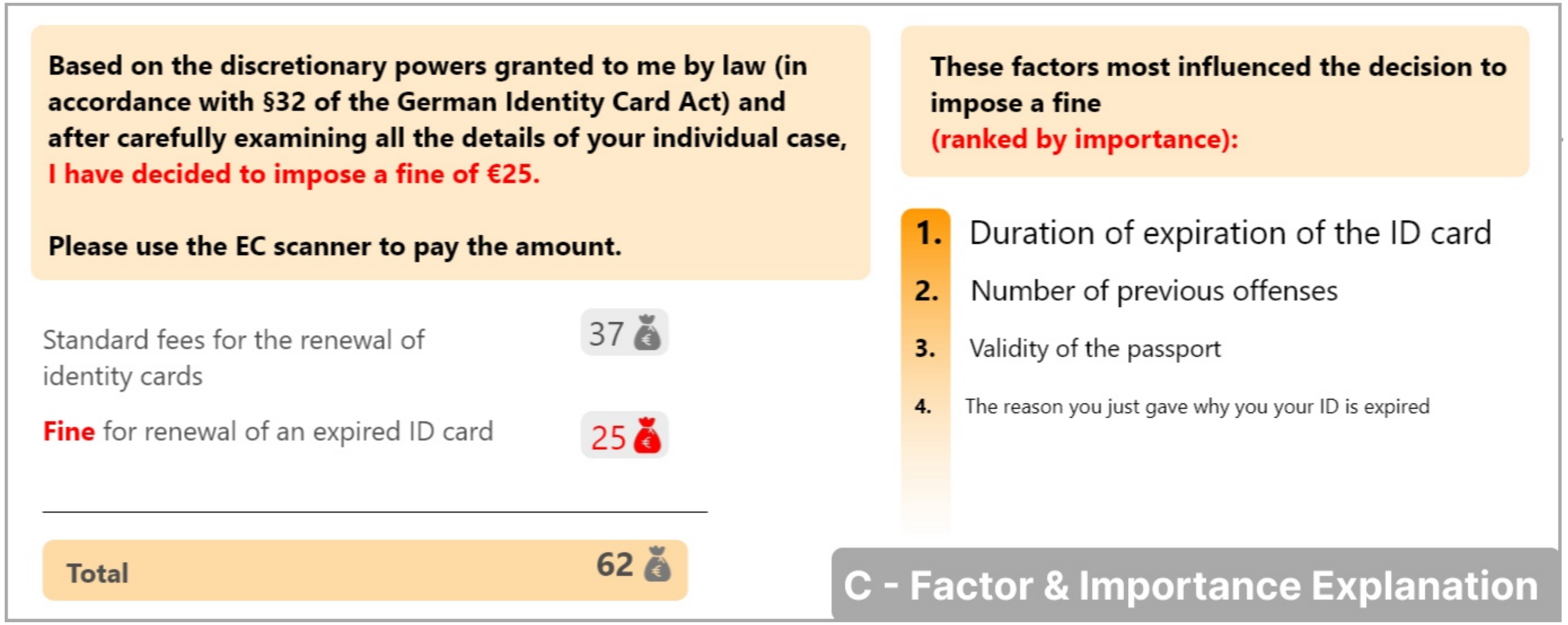
- Duration of expiration of the ID card
- Number of previous offenses
- Validity of the passport
- The reason you just gave why you your ID is expired

B - Factor Explanation

Aljuneidi et al., Why the Fine, AI?
The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations, CHI ’24

CHI '24

Explanation Level



Based on the discretionary powers granted to me by law (in accordance with §32 of the German Identity Card Act) and after carefully examining all the details of your individual case, I have decided to impose a fine of €25.

Please use the EC scanner to pay the amount.

Standard fees for the renewal of identity cards	37
Fine for renewal of an expired ID card	25
Total	**62**

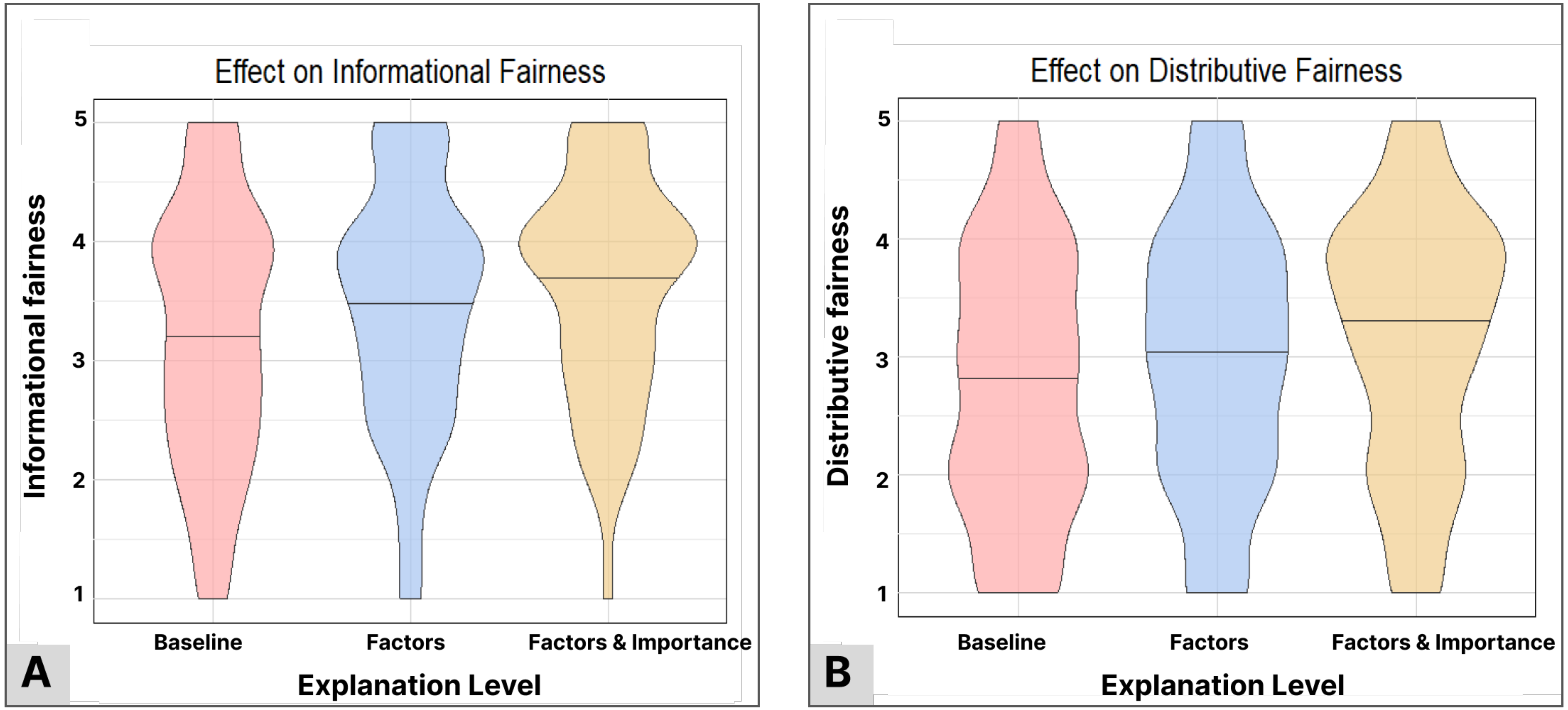
These factors most influenced the decision to impose a fine (ranked by importance):

1. Duration of expiration of the ID card
2. Number of previous offenses
3. Validity of the passport
4. The reason you just gave why you your ID is expired

C - Factor & Importance Explanation

Aljuneidi et al., Why the Fine, AI?
The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations, CHI '24

Explanation Level



Positive correlation between the level of detail provided in decision explanations and citizens' perceptions of both informational and distributive fairness

Aljuneidi et al., Why the Fine, AI?
The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations, CHI '24

CHI ’24

Explanation Level

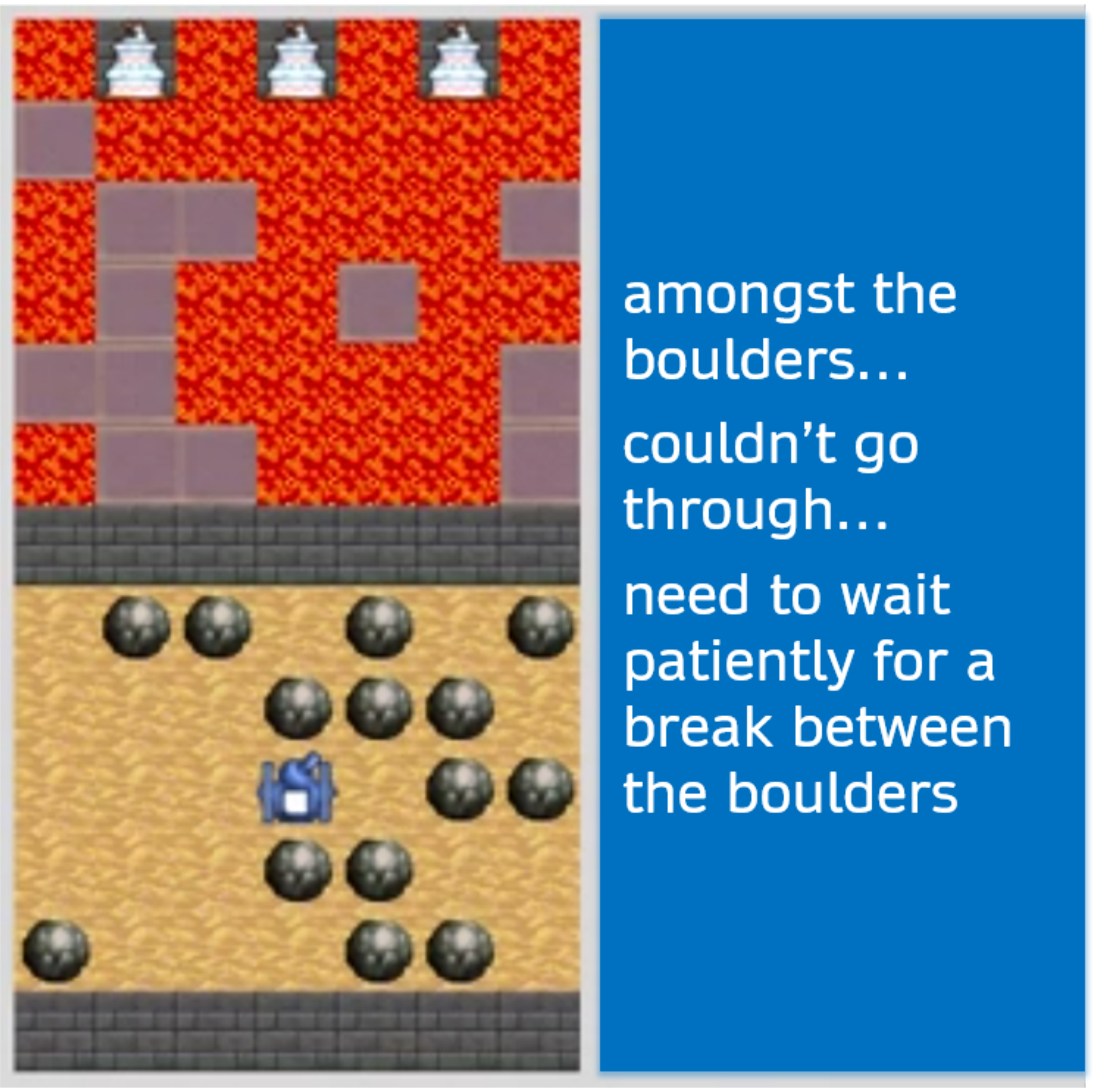
P1200: “This is a good example of a place where AI is not suitable: when decisions have to be made that afect people. I hope that such things are never introduced.”

P230: “When it comes to discretion, the AI is of course very factual, in contrast to humans, who might have turned a blind eye in the case. That makes it fairer, because decisions cannot be made based on sympathy, but it also takes away the human component.”

Aljuneidi et al., Why the Fine, AI?
The Effect of Explanation Level on Citizens' Fairness Perception of AI-based Discretion in Public Administrations, CHI ’24

CHI ’24

Feedback Adapted to User’s Background

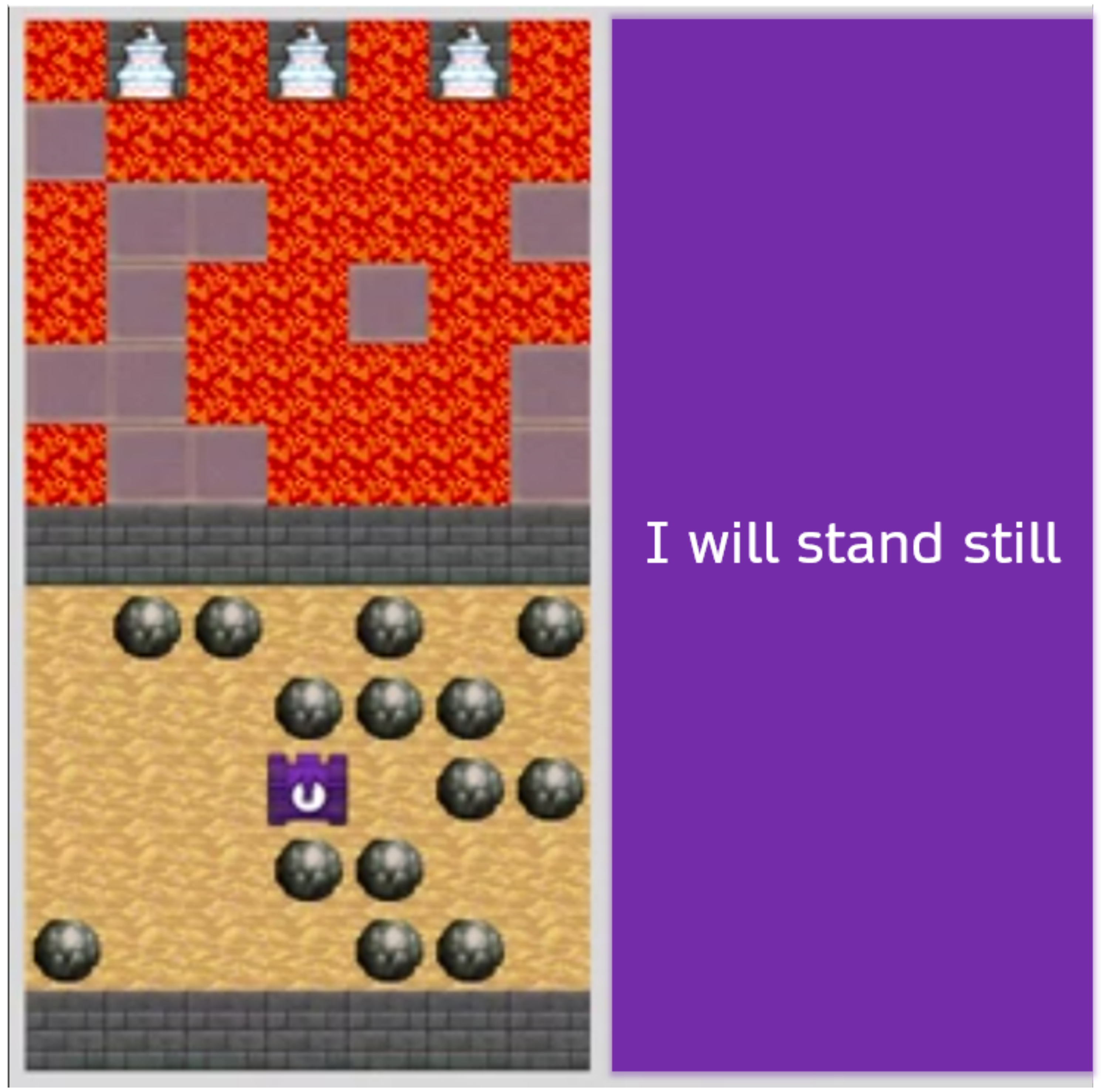


amongst the boulders…

couldn’t go through…

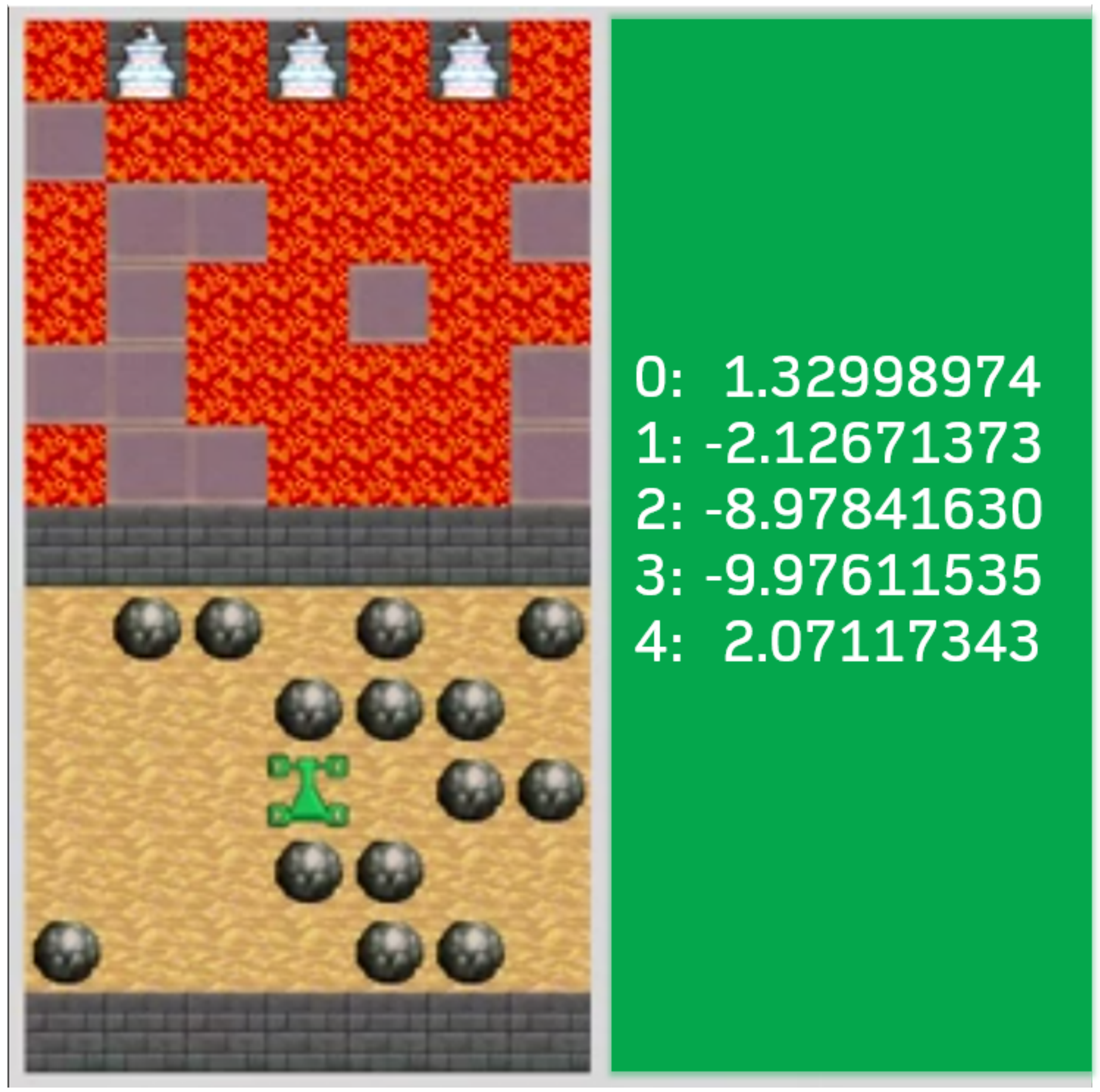
need to wait patiently for a break between the boulders

Rationale-Generating (RG)



I will stand still

Action-Declaring (AD)



```
0:  1.32998974
1: -2.12671373
2: -8.97841630
3: -9.97611535
4:  2.07117343
```

Numerical-Reasoning (NR)

Ehsan et al., The Who in XAI: How AI Background Shapes Perceptions of AI Explanations, CHI ’24

CHI ’24

Feedback Adapted to User’s Background

Unwarranted Faith in Numbers

“With [the NR robot], while I did not understand its methodology, I could see that it was using some mathematical calculations to determine which way to move. [. . .] With [the AD robot], I could not see any methodology or signs of decision-making” (A23).

Since the NR robot was “communicating in a numerical l guage that’s too hard to understand”, the numbers had a "mystery and aura of higher intelligence" (NA22, NA33). The “language of numbers”, because of its “cryptic incomprehensibility,” signalled higher-order thinking (NA6, NA1): “Because I could not understand [NR’s] output, I deemed it to be intelligent” (NA30, emphasis added).

Ehsan et al., The Who in XAI: How AI Background Shapes Perceptions of AI Explanations, CHI ’24

Feedback Adapted to User's Background

Unanticipated Explanatory Value

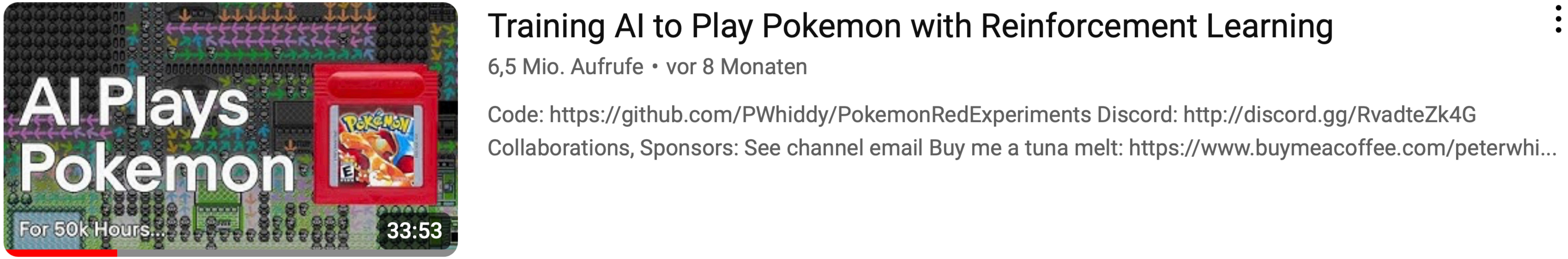
It showed that "[AD] is consistent and nothing crazy is going on where it says it went right but in actuality, it went down" (NA34).

Its "brief," "un-embellished," and "easier to understand" language that got "straight to the point" boosted its understandability (NA14, NA23, NA28, NA7). In fact, AD's "just the facts" (NA38)

AI group members perceived NR's numbers to have more explanatory value simply because they felt they could do "more with numbers" (A30) in "cases of failure and troubleshooting"(A78).

For Intelligence, NR's numerical representation and "exact values" made it appear more "valuable" than AD's "inert" statements (A23, A51, A42)

Ehsan et al., The Who in XAI: How AI Background Shapes Perceptions of AI Explanations, CHI '24



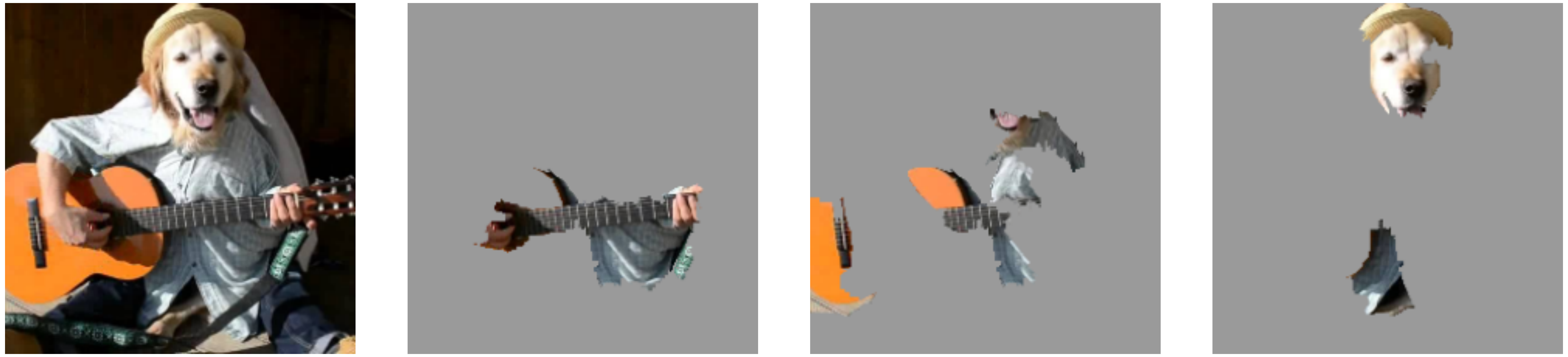
Training AI to Play Pokemon with Reinforcement Learning

6,5 Mio. Aufrufe • vor 8 Monaten

Code: https://github.com/PWhiddy/PokemonRedExperiments Discord: http://discord.gg/RvadteZk4G Collaborations, Sponsors: See channel email Buy me a tuna melt: https://www.buymeacoffee.com/peterwhi...

https://youtu.be/DcYLT37ImBY

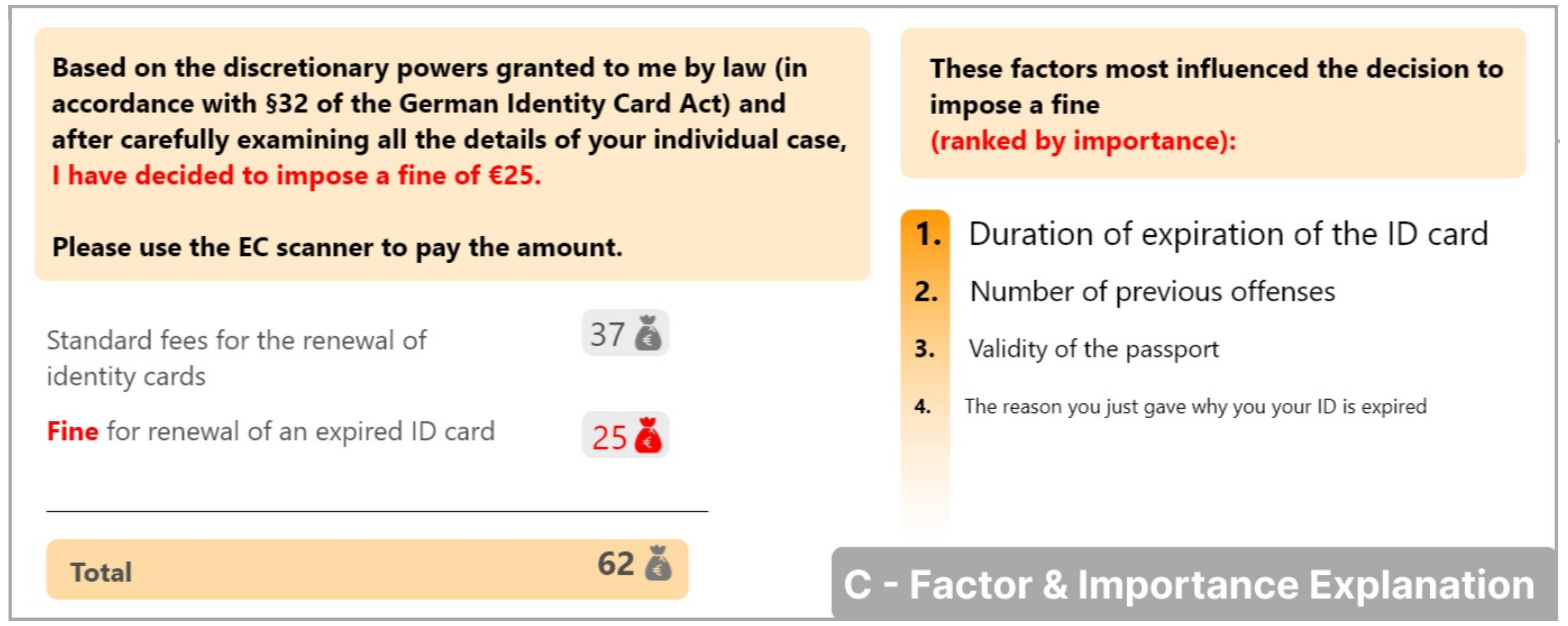
Summary



Electric Guitar

Acoustic Guitar

Labrador



Based on the discretionary powers granted to me by law (in accordance with §32 of the German Identity Card Act) and after carefully examining all the details of your individual case,
I have decided to impose a fine of €25.

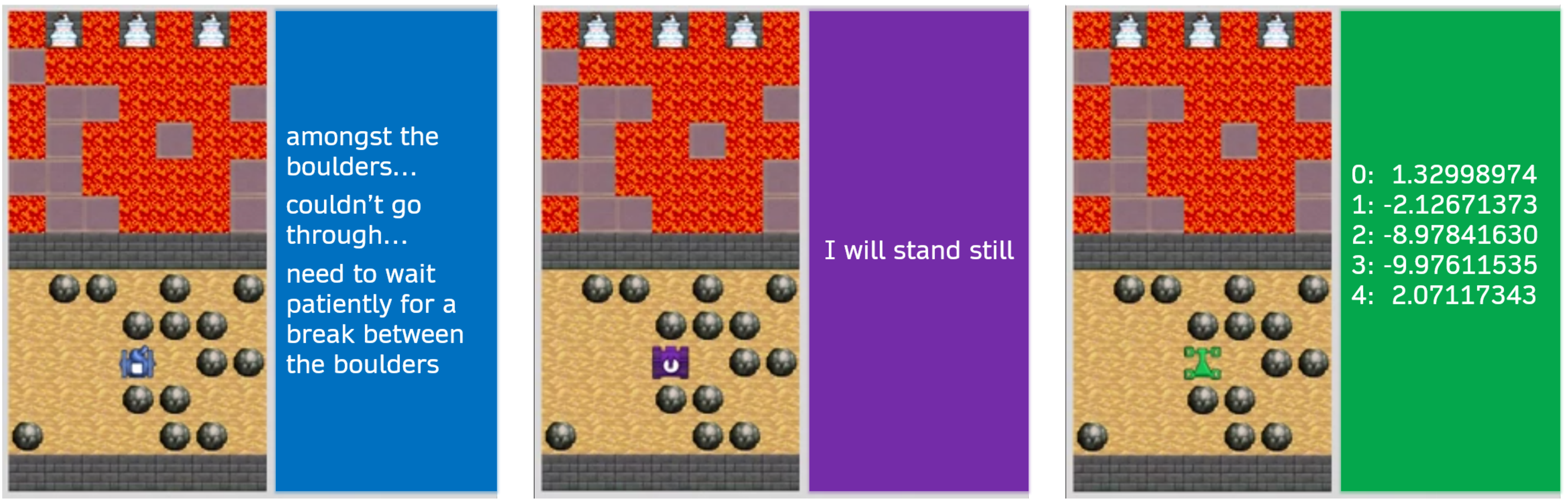
Please use the EC scanner to pay the amount.

Standard fees for the renewal of identity cards	37
Fine for renewal of an expired ID card	25
Total	**62**

These factors most influenced the decision to impose a fine
(ranked by importance):

1. Duration of expiration of the ID card
2. Number of previous offenses
3. Validity of the passport
4. The reason you just gave why you your ID is expired

C - Factor & Importance Explanation



amongst the boulders...
couldn't go through...
need to wait patiently for a break between the boulders

I will stand still

```
0:  1.32998974
1: -2.12671373
2: -8.97841630
3: -9.97611535
4:  2.07117343
```

Rationale-Generating (RG)

Action-Declaring (AD)

Numerical-Reasoning (NR)