Video presentation available here:
https://www.youtube.com/watch?v=B8xMNgIg7FI

Move to the next slide for the full slide presentation.

Thinking Like a Developer? Comparing the Attention of Humans with Neural Models of Code

Matteo Paltenghi and Michael Pradel

Software Lab, University of Stuttgart, Germany

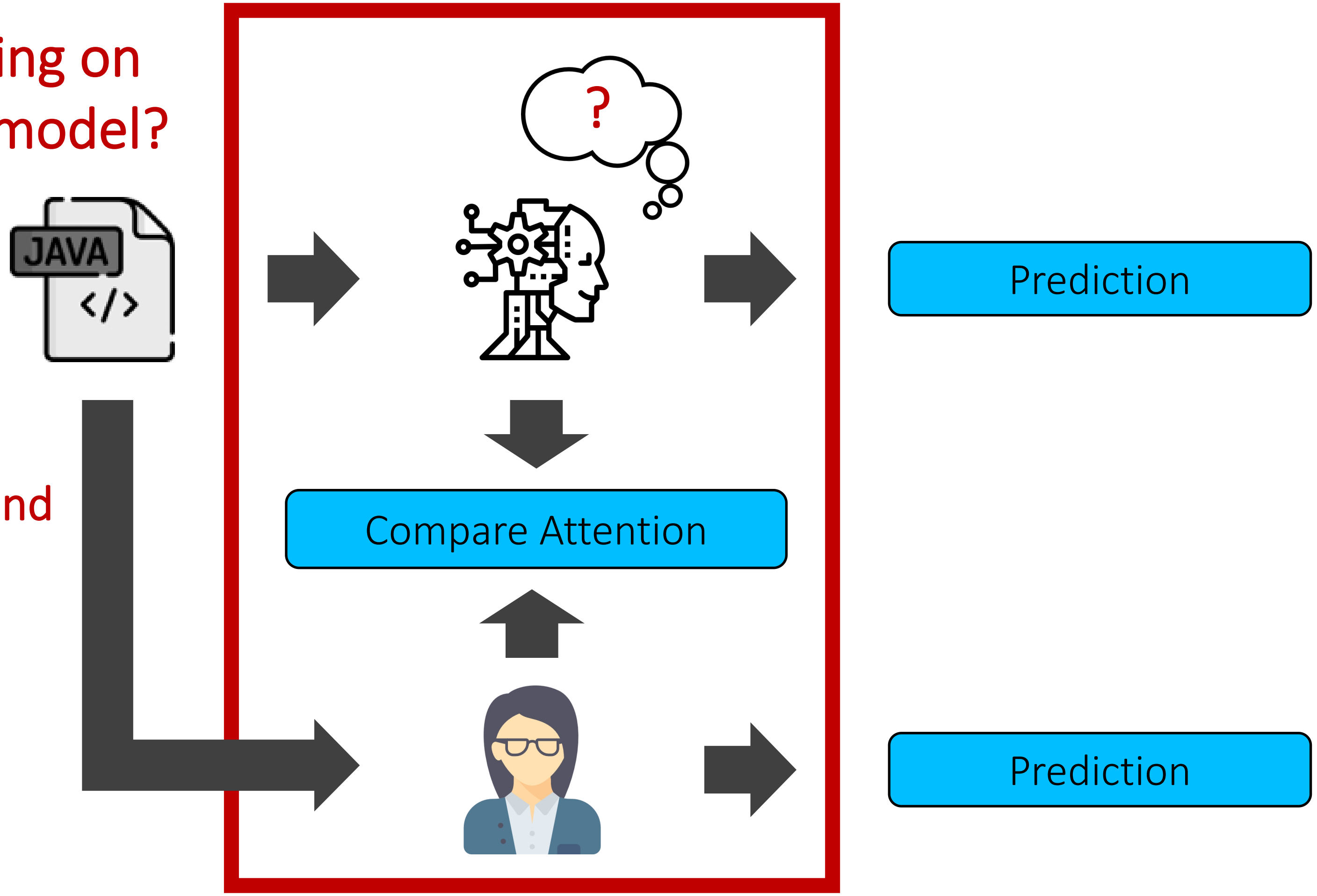
The 36th IEEE/ACM International Conference on **Automated Software Engineering** (ASE '21)

1. Motivation

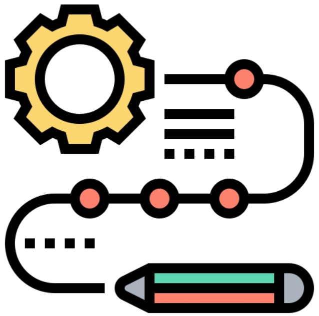
Evaluation of Neural Models of Code

- Risk: deploying a model which is **right for the wrong reason** (aka spurious dataset correlations)

What is going on inside the model?



- Our work: compare human and neural model attention
- Goal: get insights into model weaknesses

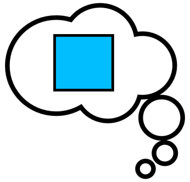


2. Methodology

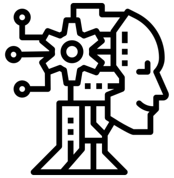
Methodology

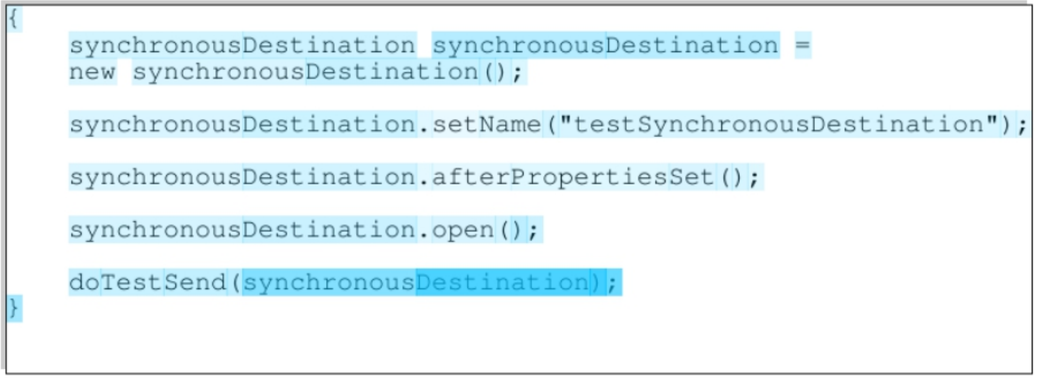
Attention Capturing

- Capture **token-level attention maps** from neural models and humans.

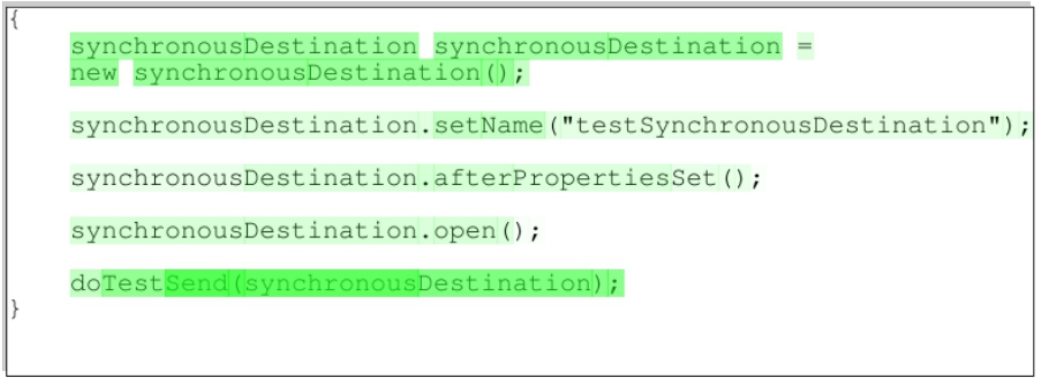


Attention Layers





```
{
    synchronousDestination synchronousDestination =
    new synchronousDestination();
    synchronousDestination.setName("testSynchronousDestination");
    synchronousDestination.afterPropertiesSet();
    synchronousDestination.open();
    doTestSend(synchronousDestination);
}
```



```
{
    synchronousDestination synchronousDestination =
    new synchronousDestination();
    synchronousDestination.setName("testSynchronousDestination");
    synchronousDestination.afterPropertiesSet();
    synchronousDestination.open();
    doTestSend(synchronousDestination);
}
```

Human Reasoning Recorder (HRR)

* darker color --> higher attention

Task Choice: Code Summarization

Method Name
[OUTPUT]

```
void sort(int arr[], int l, int r)
```

Method Body
[INPUT]

```
{
    if (l < r) {

        int m =l+ (r-l)/2;


        sort(arr, l, m);
        sort(arr, m + 1, r);


        merge(arr, l, m, r);
    }
}
```

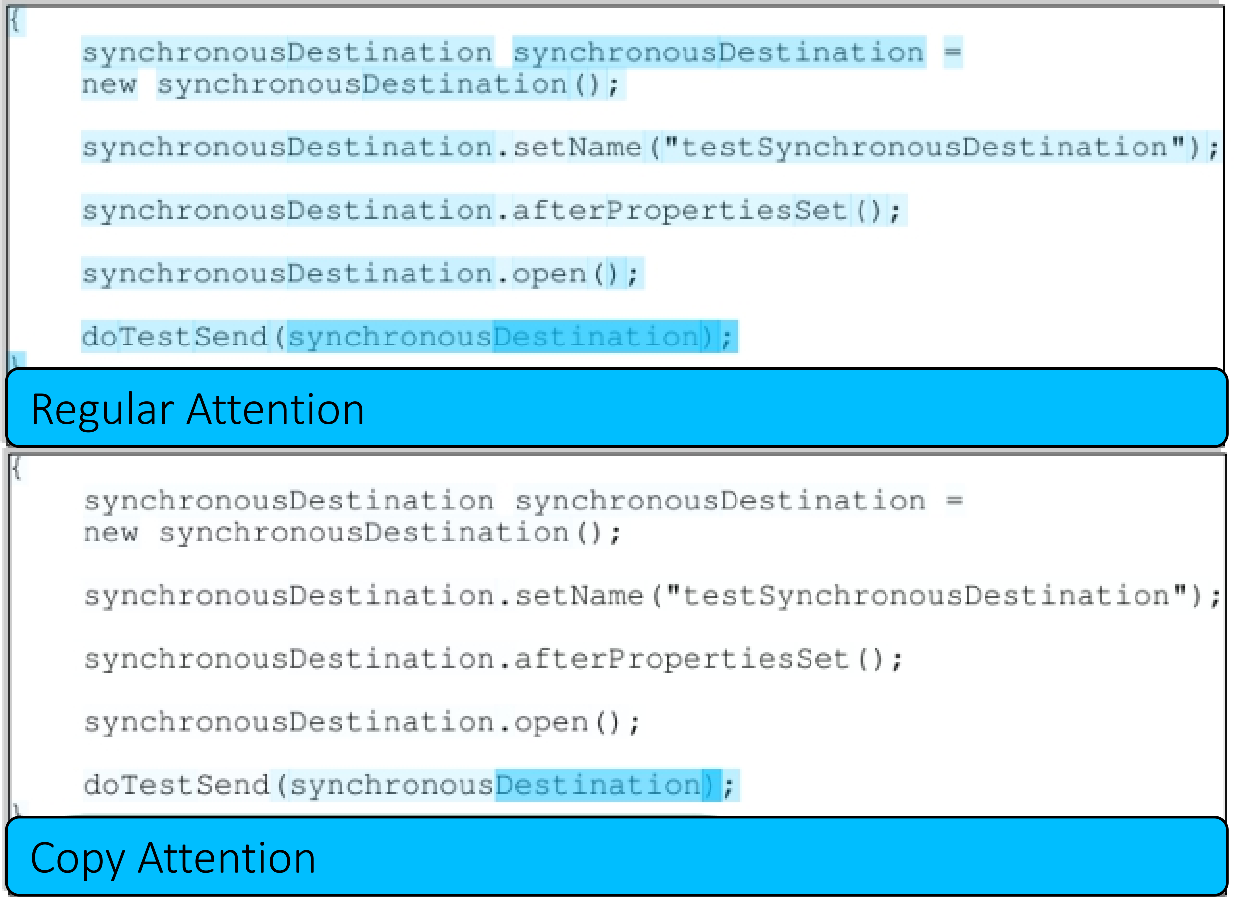
- Motivation:
  - Research interest: popularity of the task among neural models of code
  - Complex reasoning: a deeper understanding of the code is needed to name a method
- Study different model architectures:
  1. Convolutional Attention (Allamanis et al., ICML 2016)
  2. Transformer-based (Ahmad et al., ACL 2020)

Attention of Neural Models

The studied models have two types of attention:

1. **Regular attention**
2. **Copy attention** to copy verbatim tokens from the method body

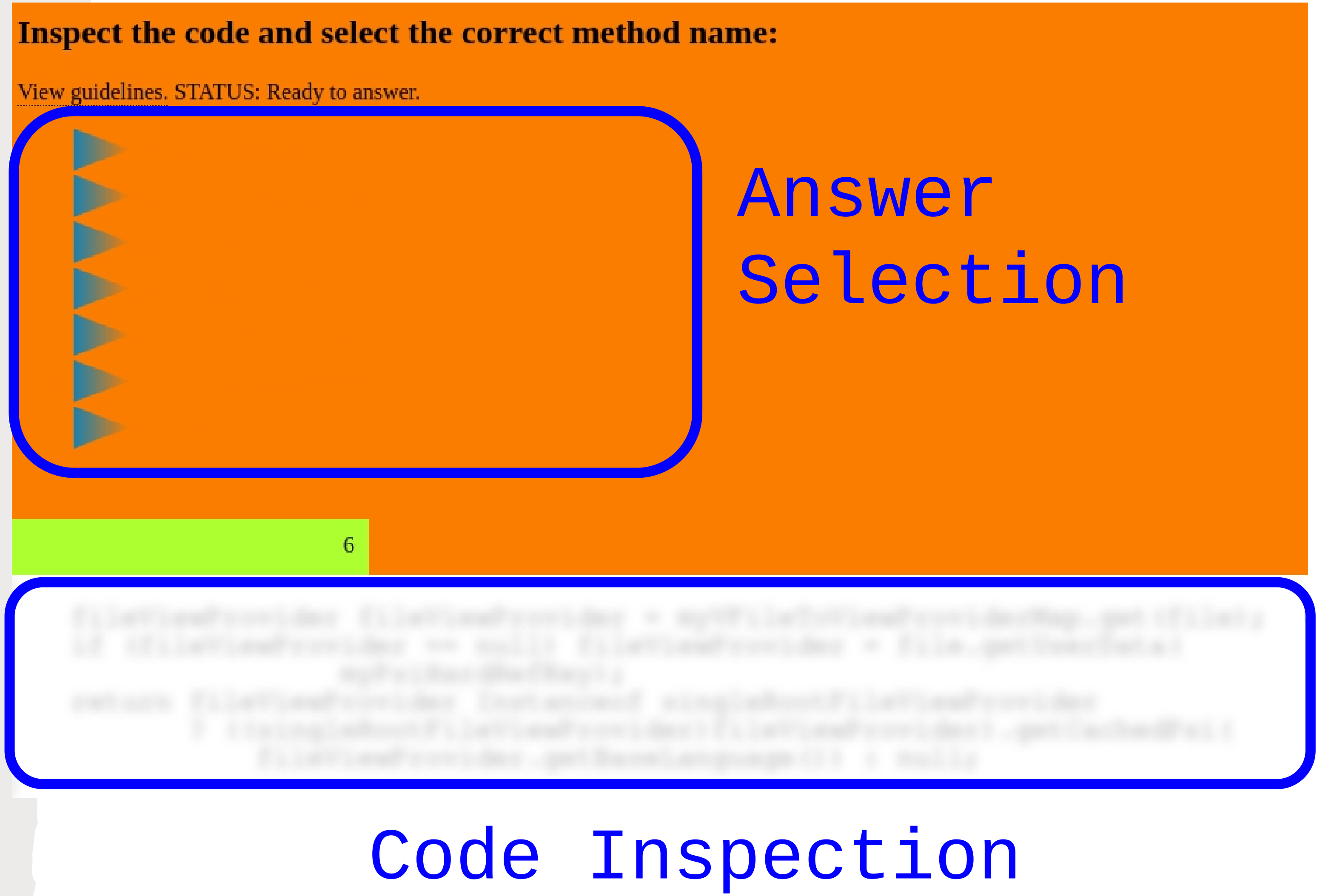
Model Prediction: testDestination()



Experimental Setup:

Human Reasoning Recorder

- **Human Task**
  choose the correct method name among 7 alternatives



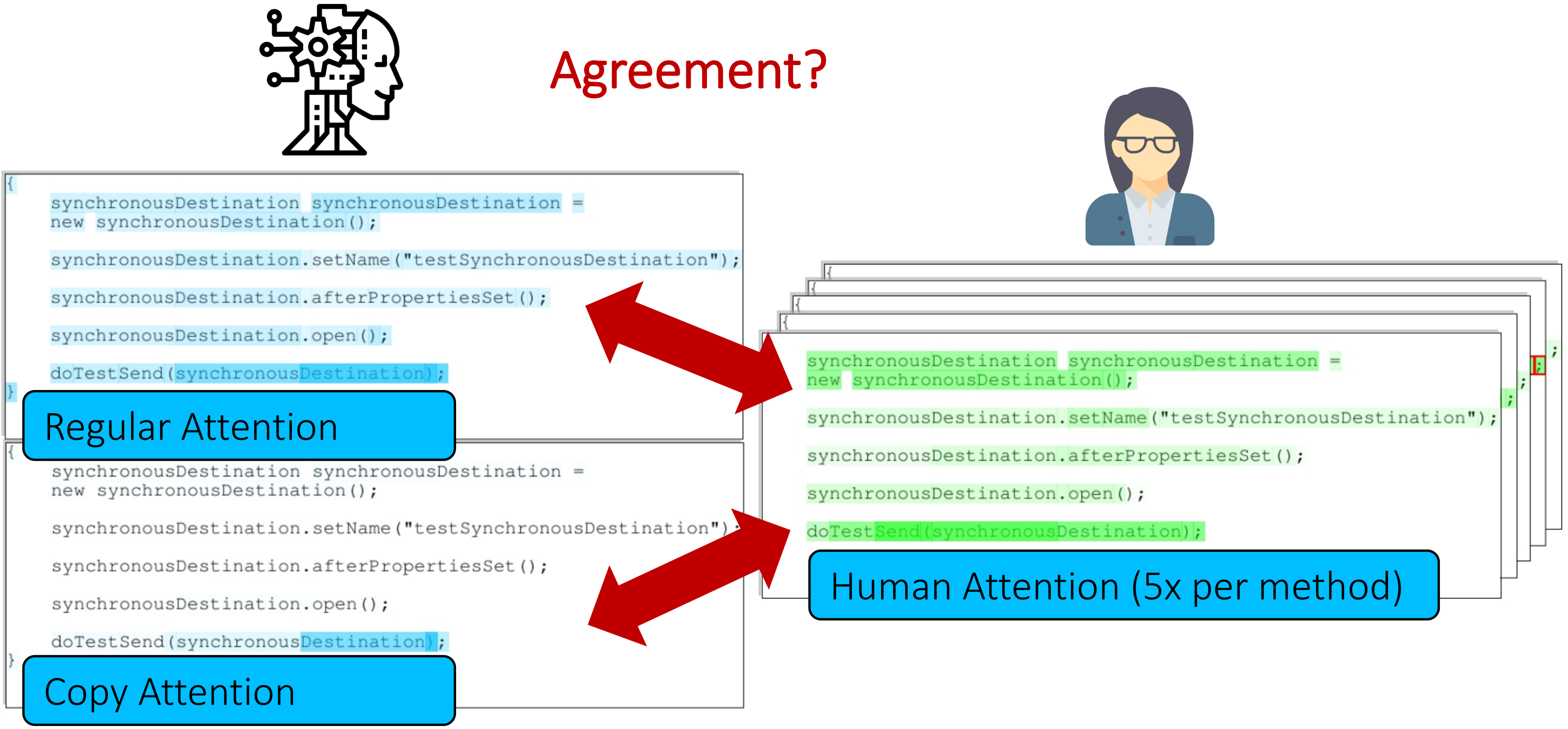
- **Fixation Time Assumption**
  The more time you stare at a token the more attention it receives

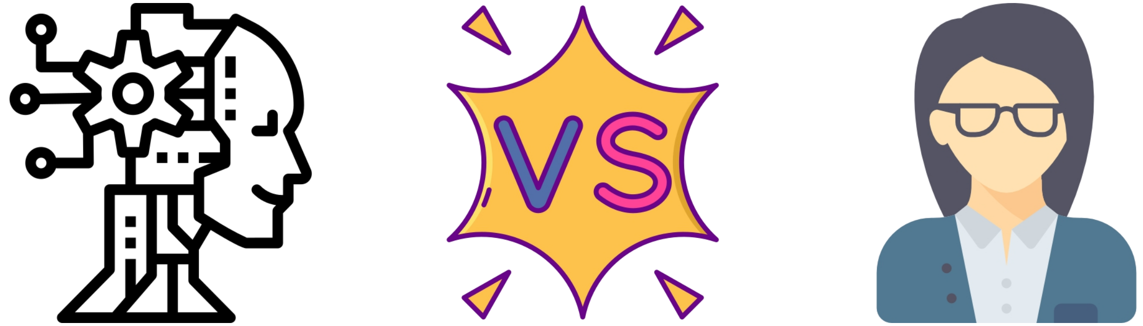
Human-Model Agreement

How to measure it?
Via **Spearman Rank Coefficient**

We compute the agreement **for each pair**:

- (Neural Model, Human)





3. Results

Human Attention Dataset

Our dataset contains:

- 1,508 human attention maps
- Methods from 10 Java Projects
- 91 participants:
  - 26 computer science students
  - 65 recruited via Amazon Mechanical Turk

Human Attention

```
log.debug("Requesting new token");
int status = getHttpClient().executeMethod(method);
if (status != 200)
{
    throw new exception("Error logging in: " + method.getS
}
document document = new saxBuilder(false).build(method.get
xPath path = xPath.newInstance("/response/token");
element result = (element)path.selectSingleNode(document);
if (result == null)
{
    element error = (element)xPath.newInstance("/response/
        document);
    throw new exception(error == null ? "Error logging in"
}
myToken = result.getTextTrim();
```

Research Question 1:

Human-model agreement?

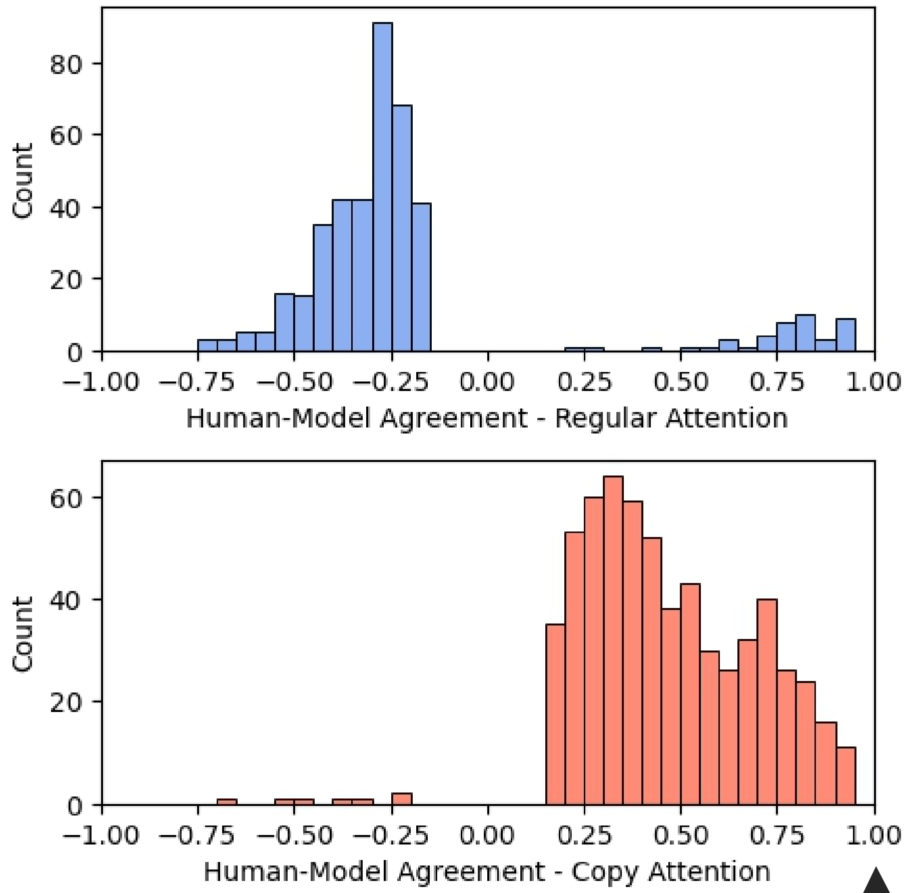
We compare each pair of human vs machine attention.

Regular attention shows a poor agreement.

Copy attention agrees with the humans.

Our work gives an **empirical justification** to the use of **copy attention, as something in agreement with the humans**.

Perfect dis-agreement



Perfect agreement

* Here you see the transformer-based model
(similar behavior for the CNN-based)

Research Question 2:

How interesting are the various kinds of token?

We quantify how much attention certain kind of tokens get w.r.t. the uniform attention scenario.

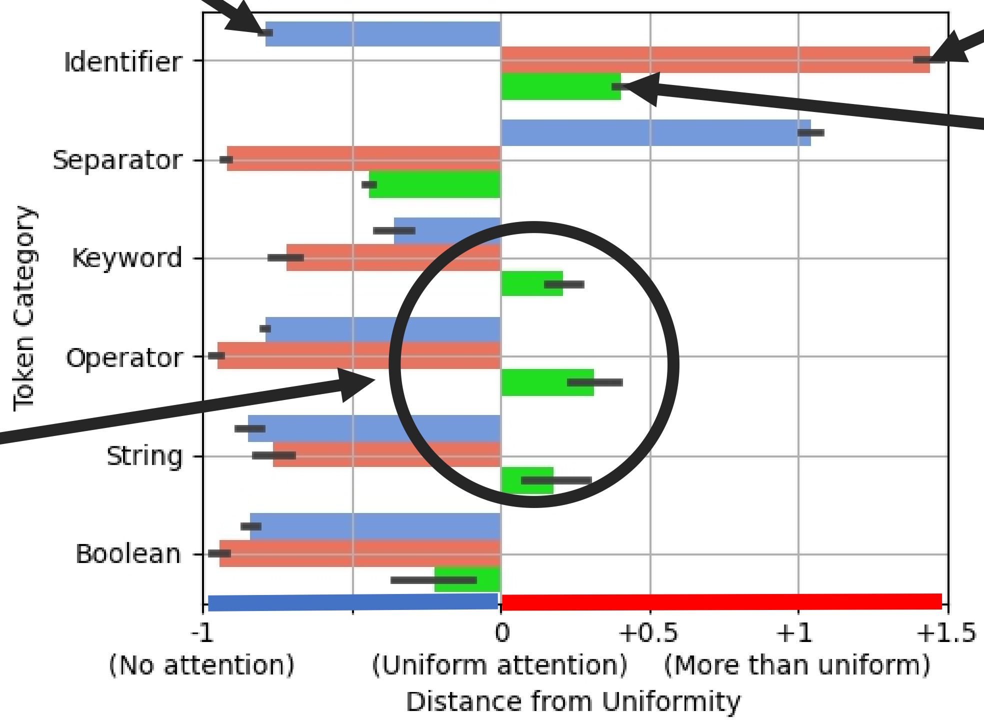
Strings, **keywords**, and **operators** are often overlooked by the models, whereas the humans give more attention to them.

Future **human-inspired neural models** should pay **more attention** to **strings, keywords, and operators**.

Regular Attention

Copy Attention

Human Attention



Less than uniform attention

Perfectly **uniform** attention

More than uniform attention

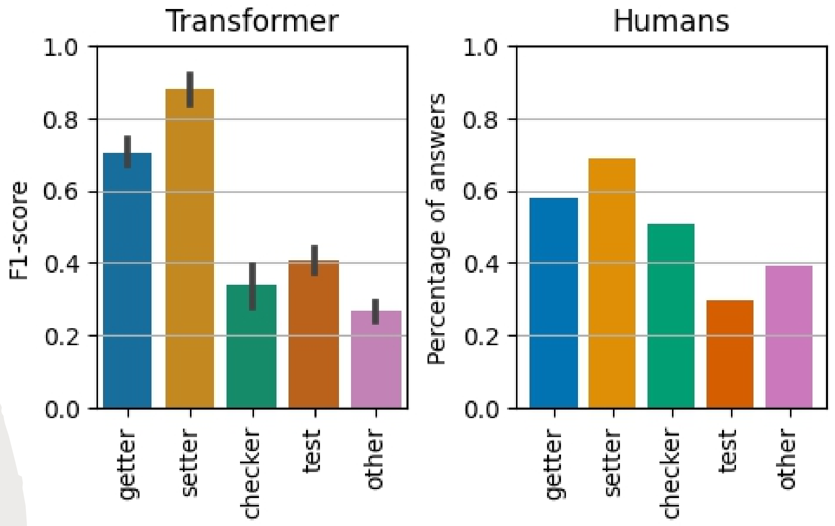
Research Question 3:

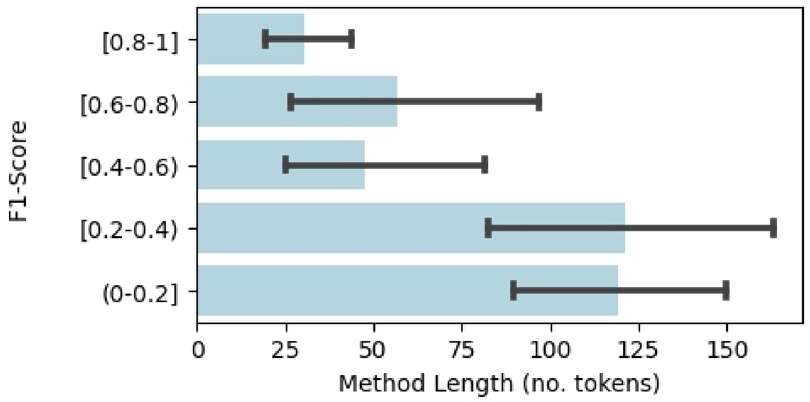
Where do humans and models **struggle** the most?

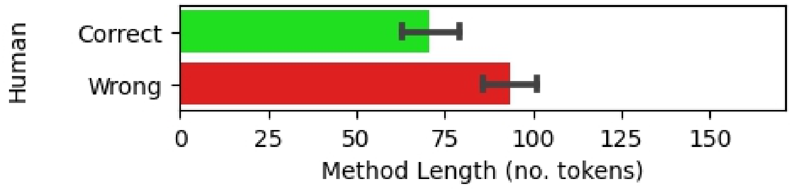
We analyze the human and model **performance** on methods of:

- different **families** (e.g., *getter, setter, test,* etc.);
- increasing **length**.

Future training datasets should include a **larger portion of "difficult" examples** for a more effective training, or different sub-datasets of increasing difficulty.







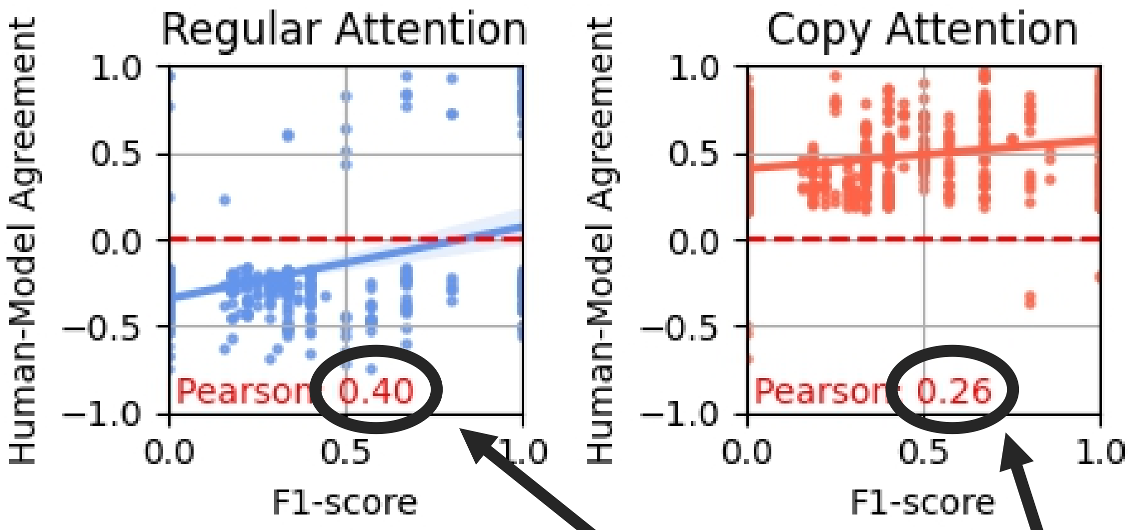
Neural models **struggle** on more challenging methods (**checkers** and **test**).

Longer methods are **harder** to summarize, both for models and humans.

Research Question 4:

Relationship between **Human-Model agreement** and **model effectiveness**?

We compute the correlation between agreement and performance with a **Pearson correlation** coefficient.



Creating **models that more closely mimic the human attention** seems a promising way toward more effective models, e.g., by using **human attention traces during training**.

A **higher human-model correlation** coincides with **more effective predictions** by the neural models.

Impact on Future Work

Ideas and Guidelines

Our work gives an **empirical justification** to the use of **copy attention, as something in agreement with the humans**.

Future **human-inspired neural models** should pay **more attention** to **strings, keywords, and operators**.

Future training datasets should include a **larger portion of "difficult" examples**.

Creating **models that more closely mimic the human attention,** seems a promising way toward more effective models.

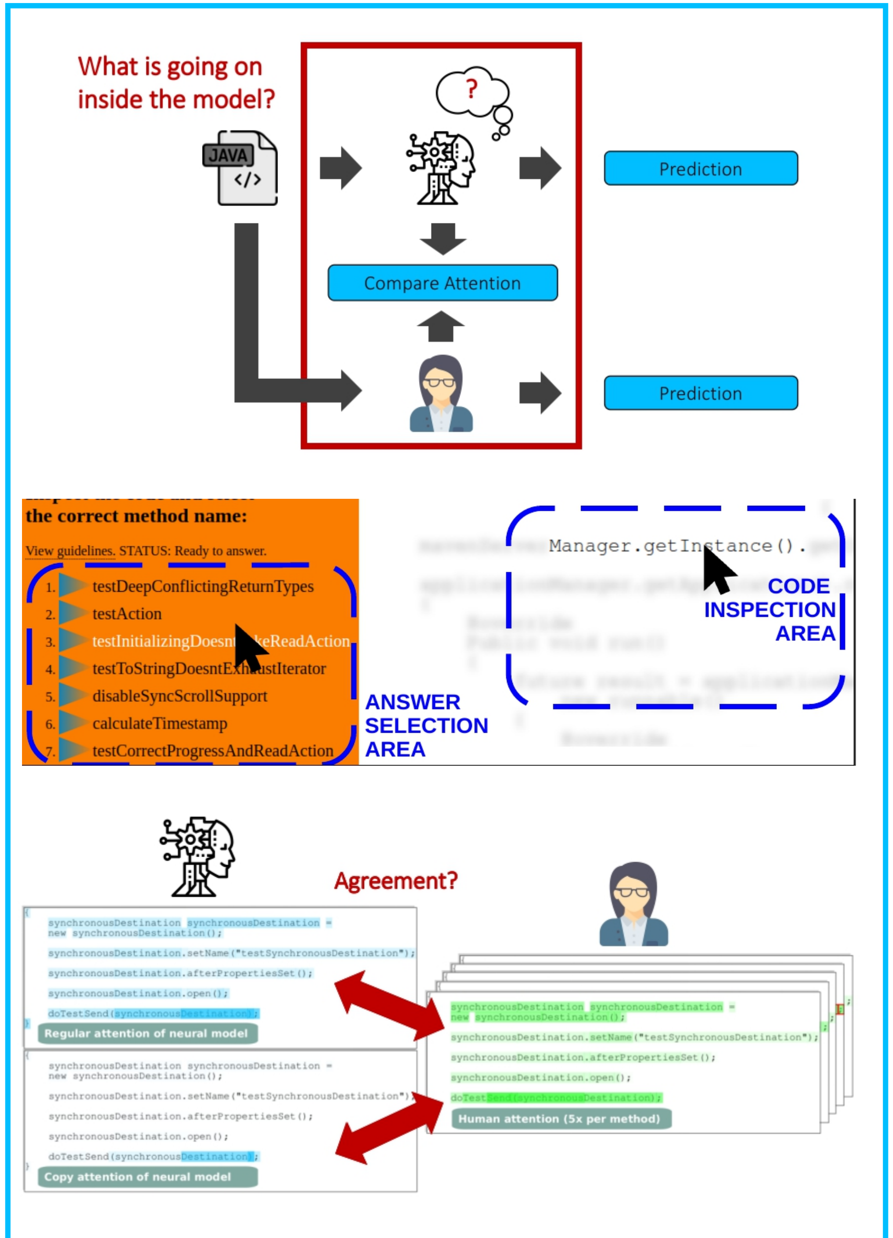
Artifacts Available

Dataset of human attention traces:

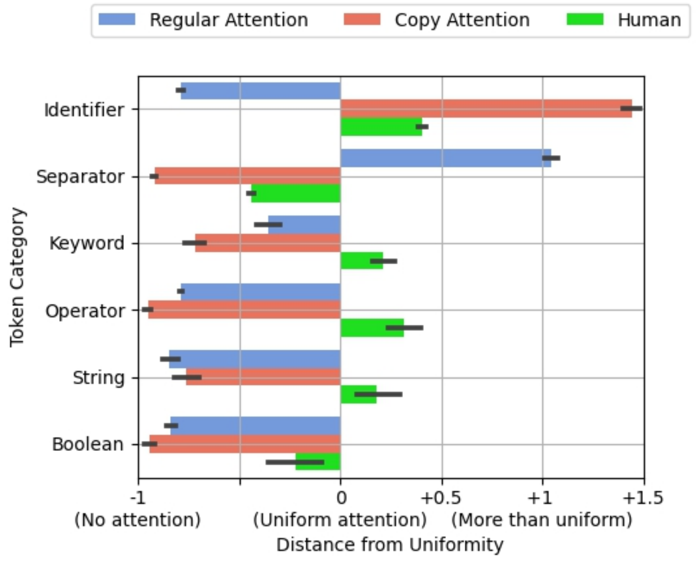
1. Benchmark another Explainable AI method.
2. Train your neural model on our human attention traces.

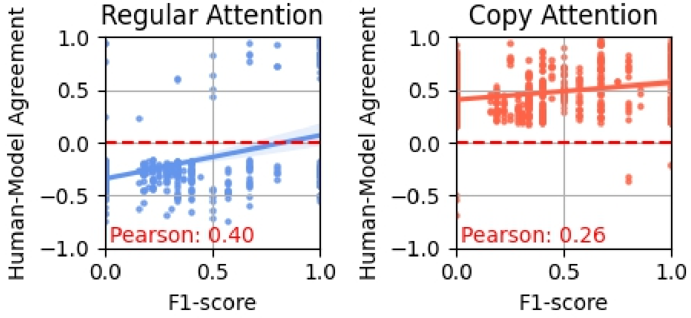
Human Reasoning Recorder:

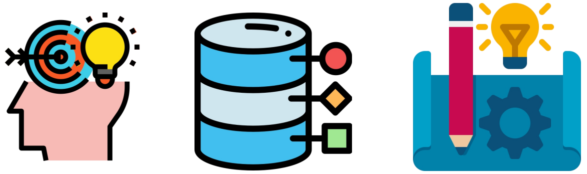
3. Use it for future human studies on source code with remote participants.



This presentation has been designed using resources from Flaticon.com







Thinking Like a Developer? Comparing the Attention of Humans with Neural Models of Code

Matteo Paltenghi and Michael Pradel
Software Lab, University of Stuttgart, Germany

Contact: mattepalte@live.it

Project: github.com/MattePalte/thinking-like-a-developer

Video presentation available here:
https://www.youtube.com/watch?v=B8xMNgIg7FI

Thanks in advance for leaving a like to the video, a simple like helps to amplify the impact of this work. Thanks! I wish you a happy and productive day.