The plan for this project

The Problem

AI on itself shouldn't be trusted but how can we generate trust for the user base that we want to support. With searching for news there are general news websites that a person trusts but how can a bot look at a website of these news sites and be sure it can be trusted.

I will dive now in some examples of news websites because this is one of the things that should be trusted in general.

The onion for example is a well-known news website that delivers satirical news posts. But still people fall for it see the following reddit subreddit r/AteTheOnion

But that was satirical now there are some actual fake news websites that want to spread news that cannot be trusted an example being Ireland Top News that should be using generative Ai for their news posts following Wikipedia. So far, I investigated it, it seemed fine and couldn't see anything that looks false.

So, I want to see how far I can go with myself to the finish line of making a seemingly trustworthy bot that people without much information on how it works can trust in the basic sense.

Certain steps are needed for this so here below you will see some activities that I can think of how to generate as much trust as possible without info dumping somebody so, they decide if it's trustworthy for them. Because with too much it will be bad

Activities

The big part is: How can a platform with actors be trusted in itself and also for newcomers to the platform.

1. How trust in AI is done.
2. What are good methods so the user also knows how it will be trusted.
3. How security across HTTP can ensured.
4. How can we be up-to-date with websites that are fake or not to be trusted.
5. How can sources that we take from be trusted.
6. Can it be used with information that is sensitive.

Links to specific activity plans

1. TrustInAI
2.
3.

4.
5.
6.

What Research method I'm going to use

- Available product analysis
  - I want to find if there are already products that already thought of these things about trust
  - 1 week.
- Best good and bad practices
  - I need to find the best and the worst of everything so further one with the card of `multi-criteria decision making` I will have enough information to make a wise decision.
  - 2 weeks.
- Community research
  - What does the community think about security, trust issues and how should they be overcome.
  - 3 days.
- SWOT analysis
  - For the reason of extending the card of `Best good and bad practices` so you will get the full picture.
  - 4 days.
- A/B testing
  - With various products there are minor changes in between with this card we can make sure if we are taking the right direction or which direction we should go into.
  - 2 days.
- Data analytics
  - With these actors there will be different datasets that will be used we need to see if it's correctly used, or we need to find another dataset. Furthermore, we can analyze it and find if there is a certain bias inside of it.
  - 1 day.
- Security test
  - Find all the risks that could be associated with the products that will be delivered and which vulnerabilities we prioritize.
  - 1 week.
- Guideline conformity analysis
  - Do I still follow the GDPR to the full extend and what about APPI for the Japanese. If this becomes global, we also need to make sure that it follows those rules also to some extend or restrict access.
  - 1 day.
- System test
  - See if everything that relates to the system is working to the extend it needs to be.
  - 3 days.
- Unit test
  - Does the actors work always what about if there is some unexpected data introduced. This is also important to know if it always works with new and

unknown data.
    - 1 week.
- Benchmark test
    - Is the product not too slow or can't it take large quantity of requests. Doesn't it too many resources.
    - 4 days.
- Ethical check
    - Is there no certain bias and how we do minimize it. What's the culture of the products and how can we satisfy as much people as possible.
    - 2 days.
- Product review
    - Does it work. Is there not something wrong with it or do we need to change something. What about the GDPR or APPI rules?
    - 1 day.
- Gap analysis
    - What's ideal solution, how close can we come to it. What does it cost (money, time and resources) to get to it. Will we change the ideal solution.
    - 1 day
- Multi-criteria decision making
    - What can be the "best" actor in this project and how did we decide that.
    - 2 days.
- Prototyping
    - Make a prototype see what it brings to the project. What does it achieve.
    - 1.5 days per week.