Note: this is not finished and preliminary for now, feel free to check it out. It should be done by Friday.

Focus on the AI act.

3.1 & 3.2 individual assessment of risk and the elements that have to be included

3.3 individual reflection (after we discuss the individual risk we can include that here)

ONLY FOCUS ON AI ACT NOT GDPR

Poor: Document the individual part (week 3 stuff as well), mention the AI act in the legal framework

Insufficient: INDIVIDUALLY assess the risk of the AI model, why is it this level? Then discuss this with the team to see if we agree or not. What laws do we have to address.

Sufficient: Discuss how our group aligned or misaligned on the legal part (risk assessment), then explain why you had a different opinion at first. Mention which requirements we have to address and which ones we don't have to address.

Good: Explain why and why not addressing certain laws based on the level of risk of our AI.

Excellent: Show evidence of us complying with the AI act, so show how we would be transparent, show how things relate to requirements (mention personal data is removed etc)

Article 6:

Our AI system is intended to be used as a safety component (safer driving)

It does use third-party as we aren't directly related to anwb (yet)

THerefore the system is high risk

Article 7:

If the commission were to assess our AI system

A: Our intended purpose is to improve road safety by warning people about dangerous roads and choosing safer options or driving more carefully in those areas.

B: The AI system would likely be used by those who already opt to drive more safely

C: Personal data is removed for the purpose of assessing road safety.

D: The AI system only informs and doesn't make any decisions by itself. Recommendations shouldn't lead to more harm, but if users drive more carelessly due to not receiving a notification of approaching a dangerous road this could lead to harm.

E: The AI system is new so hasn't already caused harm.

F: see E

G: Opting out from outcomes is practical as the driver can still choose the road themselves.

H: The users of the AI system are the drivers, they are not put in a vulnerable position by our ML model.

I: If drivers approach a high risk road they can still opt to change their route by taking a turn somewhere so the decision is corrigible, however this might be a problem if hastily taken a corner.

J: The product is likely to improve road safety.

K:

Article 9:

1: we have to establish a risk management system, and document what we do with it.

2: identify reasonable forseeable risks that can be seen when using the system; estimate risk of using, as well as misusing the app (misuse would probably be irrelevant, unless people drive less safe in high risk zones),

We can do this by identifying false positives/negatives or false classifications of roads in our case, and try to mitigate the amount.

We will adress the likelyhood and impact of the risks

For longterm we would develop strategies to minimise the risks by keeping algorithms up to date and making sure data quality is as high as possible.

Article 10:

Required to have good quality datasets for training validation and testing

We can ensure we have this by

Having high-quality data representing road safety (ANWB dataset is representative of the data)

Data should be relevant to road conditions (The data is categorised by road)

Have diverse data to avoid bias (Data is from all over Breda)

Implement processes to ensure data integrity and secure handling (Password login to access data, columns that seem wrong aren't used)

Article 11:

Create and maintain comprehensive technical documentation

We should document our architecture, algorithms and our source in a seperate file

We should state our intended purpose (Improving road safety by classifying roads as dangerous and informing end users of these roads)

Include performance metrics of the models

Maintain documentation on improvements (markdowns on models and changes)

Article 12: Keep logs to monitor the system operations (this would only be relevant post deployment; therefore it is irrelevant for the scope of our project.

Logs of outputs are still relevant so we should keep those.

Article 13: Provide clear information on ai to users

Provide detailed instructions on how the AI works

Describe the limitations of the system

Offer a way to contact services to assist.

Article 14: Ensure human oversight is in place to mitigate risk

Humans should be able to override the decisions the model makes

Staff would have to be trained to intervene (Not relevant yet in our case, since we won't actually use the model for real time cases yet.

Continuously monitor the performance and involve humans in critical decision making (Continuous monitoring is relevant during and after deployment not before.)

Article 15: Ensure systems are accurate and secure

Test the model with different conditions

Update the model to improve performance and address new data (new data might not be relevant for the scope of our project)

Implement cybersecurity measures to protect the data (Password and vpn protection on the data)

Article 16: Implement quality managmenet for the model

Develop a procedure to ensure quality remains high

Conduct external audits to ensure compliance (we can ask fellow students or perhaps the teachers to check if they feel our model is compliant)

Make sure we all try to help each other improve by integrating feedback.

Article 17: Document the quality management system with our policies and objectives

Document the procedures for design

Meet quality management standards

Improve based on received feedback (limited in scope due to our project scope)

Article 18:

Maintain documents after placing on the market (we won't actually place it so this is irrelevant)

Update the documents (irrelevant for the same reason)

Accessible to relevant authorities (irrelevant for the same reason)

Article 19:

Automatically generate logs (Output is documented earlier, new data won't be input during the scope of the project.)

Logs must be retained for a sufficient period of time to enable monitoring. (Should we log things we can maintain them but we likely won't add new events.)

Logs must be integral to prevent tampering. (See before)

Article 20:

If a large problem arises we have the duty to pull the plug, and take appropriate measures to get the model back on track. (Since our model is more of a proof of concept than an actual product this will likely not be relevant).

Article 21:

Should a relevant authority ask for information and documentation we will provide it, as well as the automatically generated logs. This will be confidentially shared. (This is unlikely to happen, but if it will we should be ready).

Article 22:

Article 22 references launching AI systems from outside the EU in the EU, since we are in the EU this is irrelevant for us.

Article 23:

Article 23 talks about an importer, if we were to sell this product to ANWB they would be the importer therefore this is not relevant for us.

Article 24:

Article 24 talks about the distributor which would also be ANWB not us who has to verify we adhere to standards.

Article 25:

Article 25 states that everyone involved within the AI is responsible for the legal distribution of the AI which would relate to us. Other parts involve handing the AI system over the creators have to make sure the new owners have the necessary knowledge to keep working on it. This woudl be relevant only if ANWB actually wishes to take over our models.

Article 26:

Organisational oversight must be present for high-risk AI systems as well as instructions to use it, this must be done by trained individuals.

Data must remain relevant for the intended purpose, and continuously monitored

Employers must know they work with AI, and public authorities must comply with registration obligations

People affected by AI decision making should be informed of this.

All these points are also covered in other points, so it doubles as a bit of a summary too.

Article 27:

We need to describe how the AI system will be used, during which time and the frequency of usage, categories of natural persons likely to be affected (no one, it is opt in), specific harm it might have on people (people potentially driving less safe because they are in a "low risk" zone)