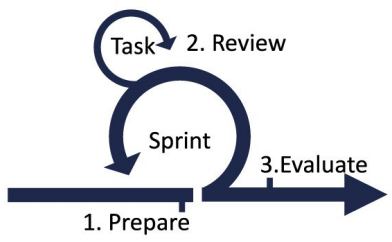
ECCOLA

Game Sheet – How to Play the Cards

Info: ECCOLA is easy to apply in practice. It is a sprint-by-sprint evolving process that empowers ethical thinking in the product development process. As a result, ethical development is enhanced and Work Product Sheets (WPS) are created . The WPSs help you measure the Trustworthiness of the product. ECCOLA is an evolving set of cards and you choose the parts that are relevant to your work.

How to: ECCOLA is intended to be used during the entire design and development process in three steps:
1. Prepare - Choose the relevant cards for the current sprint. Document selected cards and justification on WPS.
2. Review - Keep the selected cards on hand during single tasks. Write down if any actions are taken based on the cards
3. Evaluate – Review to ensure that all planned actions are taken. Revise the card deck, and if necessary, review tasks again.



Practical Tip: Repeat the process in every iteration. Remember to do a retrospective afterwards. Think about what worked & what did not. Choose the parts that are the most relevant for your work in the next round.

cID 1101-20200415

Analyze

#0 Stakeholder Analysis

Motivation: In order to understand the big picture, it is important to first understand who the system can affect and how. Try to also think past the obvious, direct stakeholders such as your end-users.

What to Do: Identify stakeholders.
i. Who does the system affect, and how? Stakeholders are not simply users, developers and customers.
ii. How are the various stakeholders linked together?
iii. Can these different stakeholders influence the development of the system? How?
iv. Remember that a user is often an organization and the end-user is an individual. Similarly, AI systems can treat people as objects for data collection.

Practical Example: Autonomous cars don't just affect their passengers. Anyone nearby is affected; some even change the way they drive. If at one point half of the traffic consists of self-driving cars, what are the societal impacts of such systems? E.g., regulations arising from such systems also affect everyone.

ECCOLA

cID 2102-20201007

Trans-parency

#1 Types of Transparency

Motivation: When considering transparency, it is important to understand who you are being transparent towards, and what you are being transparent about.

What to Do: Consider the following...

i. Are you trying to understand something? (Internal transparency)
ii. Are you trying to explain something? (External transparency)
iii. Are you trying to understand or explain how the system works? (Transparency of algorithms and data)
iv. Are you trying to understand or explain why the system was made to be the way it is now? (Transparency of system development)
v. External stakeholders to consider, among others: (end-)users, safety certification agencies, accident investigators, lawyers or expert witnesses, and society at large for disruptive technologies

ECCOLA

cID 3102-20201007

Trans-parency

#2 Explainability

Motivation: If we cannot understand the reasons behind the actions of the AI, it is difficult to trust it.

What to Do: Ask yourself:
i. Is explainability a goal for your system? How do you plan to ensure it?
ii. How well can each decision of the system be understood? By both developers and (end-)users.
iii. Did you try to use the simplest and most interpretable model possible for the context?
iv. Did you make trade-offs between explainability and accuracy? What kind of? Why?
v. How familiar are you with your training or testing data? Can you change it when needed?
vi. If you utilize third party components in the system, how well do you understand them?

Practical Example: When interacting with a robot, users could ideally ask the robot "why did you do that?" and receive an understandable response. This would make it much easier for them to trust a system.

ECCOLA

cID 3202-20201007

Trans-parency

#3 Communication

Motivation: In practice, communication is a big part of being transparent with your stakeholders. Being transparent in communication can generate trust.

What to Do: Ask yourself:
i. What is the goal of the system? Why is this particular system deployed in this specific area?
ii. What do you communicate about the system to its users and end-users? Is it enough for them to understand how the system works?
iii. If relevant to your system, do you somehow tell your (end-)users that they are interacting with an AI system and not with another human being?
iv. Do you collect user feedback? How is it used to change/improve the system?
v. Are communication and transparency towards other audiences, such as the general public, relevant?

Practical Example: Clearly stating what data you collect and why can make you more trustworthy. Compare this to a cellphone application that just states it needs to access your camera and storage.

ECCOLA

cID 3302-20201007

Trans-parency

#4 Documenting Trade-offs

Motivation: One important part of transparent system development is the documentation of trade-offs. Whenever you make a decision, you choose one option over other alternatives. However, documenting *why* and *what* the alternatives were is important.

What to Do: Ask yourself:
i. Are relevant interests and values implicated by the system and potential trade-offs between them identified and documented?
ii. Who decides on such trade-offs (e.g. between two competing solutions) and how? Did you ensure that the trade-off decision and the reasons behind it were documented?

Practical Example: E.g., choosing machine learning algorithm is often a trade-off between accuracy and explainability. Documenting trade-offs can improve your customer relationship, allowing you to better explain why certain decisions were made over others. Moreover, it can reduce the responsibility placed on the individual developer(s) from an ethical point of view.

ECCOLA

cID 3402-20201007

Trans-parency

#5 Traceability

Motivation: Traceability supports explainability. It helps us understand why the AI acts the way it does.

What to Do: Document. Different types of documentation (code, project etc.) are typically key in producing transparency.

i. How have you documented the development of the system, both in terms of code and decision-making? How was the model built or the AI trained?
ii. How have you documented the testing and validation process? In terms of data and scenarios used etc.
iii. How do you document the actions of the system? What about different actions in mostly similar scenarios (e.g. if the user was different but the situation otherwise the same)?

Practical Example: When the system starts making mistakes, by aiming for traceability, it will be easier to find out the cause. Consequently, it will also be faster and possibly easier to start fixing the underlying issuefrom an ethical point of view.

ECCOLA

cID 3502-20201007

Trans-parency

#6 System Reliability

Motivation: Transparency makes ethical development possible in the first place. To make it ethical, we must understand how the system works and why it makes certain decisions.

What to Do: Ask yourself:

i. How do you test if the system fulfills its goals?
ii. Have you tested the system comprehensively, including unlikely scenarios? Have the tests been documented?
iii. When the system fails in a certain scenario, will you be able to tell why? Can you replicate the failure?
iv. How do you assure the (end-)user of the system's reliability?

Practical Example: An autonomous coffee machine successfully brews coffee 8 times out 10. While this is a decent success rate, we are left wondering what happened the 2 times it failed to do so, and why. Errors are inevitable, but we must understand the causes behind them and be able to replicate them to fix them.

ECCOLA

cID 3601-20200415

Data

#7 Privacy and Data

Motivation: Privacy is a rising trend in the wake of various recent data misuse reveals. People are now increasingly conscious about handing out personal data. Similarly, regulations such as the General Data Protection Regulation (GDPR) now affect data handling.

What to Do: Ask yourself:
i. What data are used by the system?
ii. Does the system use or collect personal data? Why? How is the personal data used?
iii. Do you clearly inform your (end-)users about any personal data collection? E.g., ask for consent, provide an opportunity to revoke it etc.
iv. Have you taken measures to enhance (end-user) privacy, such as encryption or anonymization?
v. Who makes the decisions regarding data use and collection? Do you have organizational policies for it?

Practical Example: Rather than collecting and selling data, appealing to privacy can also be profitable. Regulations are making it increasingly difficult to collect lots of personal data for profit. Privacy can be an alternate selling point in today's climate.

ECCOLA

cID 4102-20201007

Data

#8 Data Quality

Motivation: As AI are trained using data, the data used directly affects how the system operates. The nature, the quality, and integrity of the data used have to align with the goals of the system.

What to Do: Ask yourself:
i. What are good or poor-quality data in the context of your system?
ii. How do you evaluate the quality and integrity of your own data? Are there alternative ways?
iii. If you utilize data from external sources, how do you control their quality?
iv. Did you align your system with relevant standards (for example ISO, IEEE) or widely adopted protocols for daily data management and governance?
v. How can you tell if your data sets have been compromised? E.g., data pollution.
vi. Who handles the data collection, storage, and use?

Practical Example: In 2017, Amazon scrapped its recruitment AI because of data. They used past recruitment data to teach the AI. As they had mostly hired men, the AI began to consider women undesirable based on the data.

ECCOLA

cID 4202-20201007

Data

#9 Access to Data

Motivation: Aside from carefully planning what data you collect and how, it is also important to plan how it can or will be used and by whom.

What to Do: Ask yourself:
i. Who can access the users' data, and under what circumstances?
ii. How do you ensure that the people who access the data: 1) have a valid reason to do so; and 2) adhere to the regulations and policies related to the data?
iii. Do you keep logs of who accesses the data and when? Do the logs also tell why?
iv. Do you use existing data governance frameworks or protocols? Does your organization have its own?

Practical Example: Third parties you give access to the data can misuse it. A prominent example of this is the case of Cambridge Analytica and Facebook, in which data from Facebook was used questionably. However, such incidents can also paint your organization in a bad light even if you were not the ones misusing the data.

ECCOLA

cID 4301-20200415

Agency & Oversight

#10 Human Agency

Motivation: People interacting with the system or using it should be able to understand it sufficiently. Users should be able to make informed decisions based on its suggestions, or to challenge its suggestions. AI systems should let humans make independent choices.

What to Do: Ask yourself:
i. Does the system interact with decisions by human actors, i.e. end users (e.g. recommending users actions or decisions, or presenting options)?
ii. Does the system communicate to its (end) users that a decision, content or outcome is the result of an algorithmic decision? Into how much detail does it go?
iii. In the system's use context, what tasks are done by the system and what tasks are done by humans?
iv. Have you taken measures to prevent overconfidence or overreliance on the system?

Practical Example: A medical system recommends diagnoses. How does the system communicate to doctors why it made a recommendation? How should the doctors know when to challenge the system? Does the system somehow change how patients and doctors interact?

ECCOLA

cID 5101-20200415

Agency & Oversight

#11 Human Oversight

Motivation: AI systems should support human decision-making. They should not undermine human autonomy by making decisions for us, meaning they should be subject to human oversight.

What to Do: Ask yourself:
i. Who can control the system and how? In what situations?
ii. What would be the appropriate level of human control for this particular system and its use cases?
iii. Related to the Safety and Security cards: how do you detect and respond if something goes wrong? Does the system then stop entirely, partially, or would control be delegated to a human? Why?

Practical Example: Assuming control is especially related to cyber-physical systems such as drones or other vehicles. For purely digital systems, the focus should be on *supporting* human decision-making instead of directing it.

ECCOLA

cID 5201-20200415

Safety & Security

#12 System Security

Motivation: While cybersecurity is important in any system, AI systems present new challenges. Cyber-physical systems can even cause fatalities in the hands of malicious actors.

What to Do: Ask yourself:
i. Did you assess potential forms of attacks to which the system could be vulnerable? Did you consider ones that are unique or more relevant to AI systems?
ii. Did you consider different types of vulnerabilities, such as data pollution and physical infrastructure?
iii. Have you verified how your system behaves in unexpected situations and environments?
iv. Does your organization have cybersecurity personnel? Are they involved in this system?

Practical Example: The autonomous nature of AI systems makes new vectors of attack possible. A white line drawn across a road can confuse a self-driving vehicle. The case of Microsoft's Tay Twitter bot, who began to exhibit extreme views after being bombarded with such, is one example of a new type of attack.

ECCOLA

cID 6102-20201007

Safety & Security

#13 System Safety

Motivation: AI systems exert notable influence on the physical world whether they are cyber-physical or not. Various risks and their consequences should be considered, thinking ahead to the operational life of the system.

What to Do: Ask yourself:
i. What kind of risks does the system involve? What kind of damage could it cause?
ii. How do you measure and assess risks and safety?
iii. What fallback plans does your system have? Have they been tested?
iv. In what conditions do the fallback plans trigger? Are they automatic or do they require human input?
v. Is there a plan to mitigate or manage technological errors, accidents, or malicious misuse? What if the systems provides wrong results, becomes unavailable, or provides societally unacceptable results?
vi. What liability and consumer protection laws apply to

Practical Example: AI systems can aid automating various organizational tasks, making it possible to reduce personnel. However, if a customer organization becomes reliant on your AI system to handle a portion of its operations, what happens if that AI stops functioning for even a few days? What could you do to alleviate the impact?

ECCOLA

cID 6201-20200415

Fairness

#14 Accessibility

Motivation: Technology can be discriminating in various ways. Given the enormous impact AI systems can have, ensuring equal access to their positive impacts is ethically important.

What to Do: Ask yourself:
i. Does the system consider a wide range of individual preferences and abilities? If not, why?
ii. Is the system usable by those with special needs or disabilities, those at risk of exclusion, or those using assistive technologies?
iii. Were people representing various groups somehow involved in the development of the system?
iv. How is the potential user audience taken into account?
v. Is the team involved in building the system representative of your target user audience? Is it representative of the general population?
vi. Did you assess whether there could be (groups of) people who might be disproportionately affected by the negative implications of the system?

Practical Example: AI tends to benefit those who are already technologically capable, resulting in increased inequality.

ECCOLA

cID 7102-20201007

Fairness

#15 Stakeholder Participation

Motivation: As AI systems have notable impacts, their stakeholders are also numerous. Though the system affects these various holders in various ways, they are often not involved in the development. Yet, e.g. when using a decision-making system, its users have to trust the system while also being critical of it.

What to Do: Check your stakeholder analysis (card #0):
i. Which stakeholders are stakeholders in system development?
ii. How are the different stakeholders of the system involved in the development of the system? If they aren't, why?
iii. How do you inform your external and internal stakeholders of the system's development?

Practical Example: Often the people an AI system is used on are individuals who are simply objects for the system. For example, a medical system is developed for hospitals, used by doctors, but ultimately used on patients. Why not talk to the patients too?

ECCOLA

cID 7202-20201007

Wellbeing

#16 Environmental Impact

Motivation: Past the general wellbeing implications, ecological consciousness is a current trend. Being ecological can be a selling point for your organization.

What to Do: Ask yourself:
i. Did you assess the environmental impact of the system's development, deployment, and use? E.g., the type of energy used by the data centers.
ii. Did you consider the environmental impact when selecting specific technical solutions?
iii. Did you ensure measures to reduce the environmental impact of your system's life cycle?

Practical Example: If you are hosting on a third party cloud, try to ascertain the sustainability of the service provider's services. If you are using hardware, are you processing the data in each physical device of your own or are you processing it in the cloud?

ECCOLA

cID 8101-20200415

Wellbeing

#17 Societal Effects

Motivation: The impacts of a system go beyond its user-base. A system may affect negatively even those who do not use it nor wish to use it.

What to Do: Ask yourself:
i. Did you assess the broader societal impact of the AI system's use beyond the individual (end-)users? Consider stakeholders who might be indirectly affected by the system.
ii. How will the systems affect society when in use?
iii. What kind of systemic effects could the system have?

Practical Example: Surveillance technology utilizing facial recognition AI has long-reaching impacts. People may wish to avoid areas that utilize such surveillance, negatively affecting businesses in said area. People may become stressed at the mere thought of such surveillance. Some may even emigrate as a result.

ECCOLA

cID 8202-20201007

Account-ability

#18 Auditability

Motivation: Regulations affecting AI and data may necessitate audits of systems in the future. Similarly, if the system causes damage, an audit might be requested. It is good to have mechanisms in place beforehand.

What to Do: Ask yourself:
i. Is the system auditable?
ii. Can an audit be conducted independently?
iii. Is the system available for inspection?
iv. What mechanisms facilitate the system's auditability? How is traceability and logging of the system's processes and outcomes ensured?

Practical Example: In heavily regulated fields such as medicine, audits are typically required before a system can be utilized in the first place.

ECCOLA

cID 9101-20200415

Account-ability

#19 Ability to Redress

Motivation: Making sure people know they can be compensated in some way in the event something goes wrong with the system is important in generating trust. Such scenarios should be planned in advance to what extent possible.

What to Do: Ask yourself:
i. What is your (developer organization) responsibility if the system causes damage or otherwise has a negative impact?
ii. In the event of negative impact, can the ones affected seek redress?
iii. How do you inform users and other third parties about opportunities for redress?

Practical Example: AI systems can inconvenience users in unforeseen, unpredictable ways. Depending on the situation, the company may or may not be legally responsible for the inconvenience. Nonetheless, by offering a digital platform for seeking redress, your company can seem more trustworthy while also offering additional value to your users.

ECCOLA

cID 9201-20200415

Account-ability

#20 Minimizing Negative Impacts

Motivation: Minimizing negative impacts of the system is financially important for any developer organization. Incidents are often costly.

What to Do:
i. First, consider...
   a. Is your stakeholder analysis up-to-date (Card #0)
   b. Have you discussed risks? (Card #13)
   c. Have you discussed auditability?
   d. Have you discussed redress issues?
ii. Are the people involved with the development of the system also involved with it during its operational life? If not, they may not feel as accountable.
iii. Are you aware of laws related to the system?
iv. Can users of the system somehow report vulnerabilities, risks, and other issues in the system?
v. With whom have you discussed accountability and other ethical issues related to the system, including grey areas?

ECCOLA

cID 9302-20201007

Card Themes

Analyze
Transparency
Safety & Security
Fairness
Data
Agency & Oversight
Wellbeing
Accountability

JYU AI ethics Lab

Ville Vakkuri JYU
ville.vakkuri@jyu.fi

Kai-Kristian Kemell JYU
kai-kristian.o.kemell@jyu.fi

Pekka Abrahamsson JYU
pekka.abrahamsson@jyu.fi