VALUES TO METRICS TOOLKIT

AI, MEDIA & DEMOCRACY LAB
UNIVERSITEIT VAN AMSTERDAM

INTRODUCTION

You are about to start the excercise. This card provides an overview of the different steps in the process.

For additional guidance, refer to the accompanying information sheet. The **Metric Builder** can be found on the back of the information sheet and is used in **Step 3**.

STEP 1 – DEFINE OBJECTIVES

STEP 2 – PRIORITIZE

STEP 3 – DESIGN METRICS

STEP 4 – NOW, HOW, WOW

STEP 5 – NEXT STEPS

STEP 6 – WRAP UP

STEP 1

DEFINE OBJECTIVES

DEFINE OBJECTIVES

An objective is something the recommender system is trying to accomplish.

We define an objective by answering **why** the system is making recommendations and **who** it is making the recommendations for, which will help to answer **what** it is recommending. Discuss each of the three questions for each objective.

A recommender system could aim to fulfill multiple objectives, or multiple separate recommender systems could each have their own objectives.

The objectives will inform the next steps of metric design, and create more understanding within your group on what the recommender needs to do.

→ Next step: Read the instructions on the **objective card**.

STEP 1

OBJECTIVE

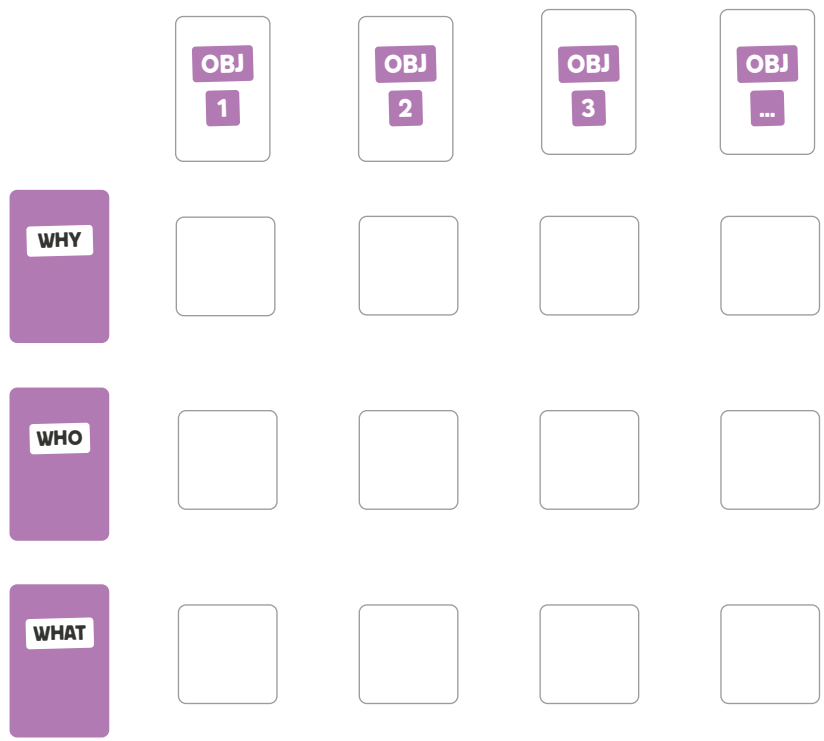
DEFINE OBJECTIVES

DEFINE OBJECTIVES

For each objective, address the questions: *Why*, *Who*, and *What*, where *Why* and *Who* will often shape *What* you do.

Create as many objectives as possible, ensuring every member's role and expertise is represented in at least some of them.

Write each answer on a post-it and place them in a grid like the one below.



STEP 1

OBJECTIVE

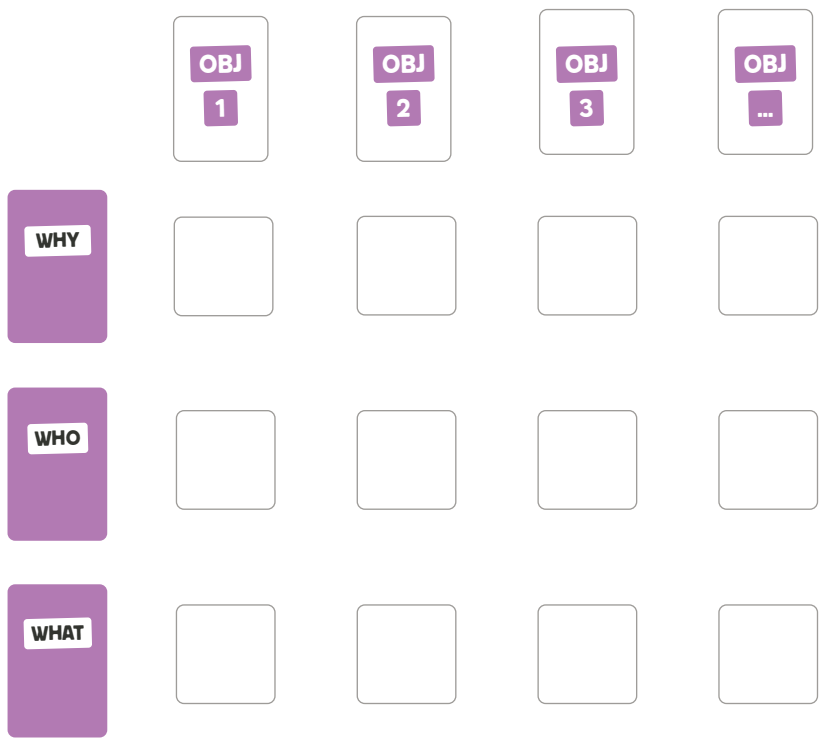
DEFINE OBJECTIVES

DEFINE OBJECTIVES

For each objective, address the questions: *Why*, *Who*, and *What*, where *Why* and *Who* will often shape *What* you do.

Create as many objectives as possible, ensuring every member's role and expertise is represented in at least some of them.

Write each answer on a post-it and place them in a grid like the one below.



STEP 1

OBJECTIVE

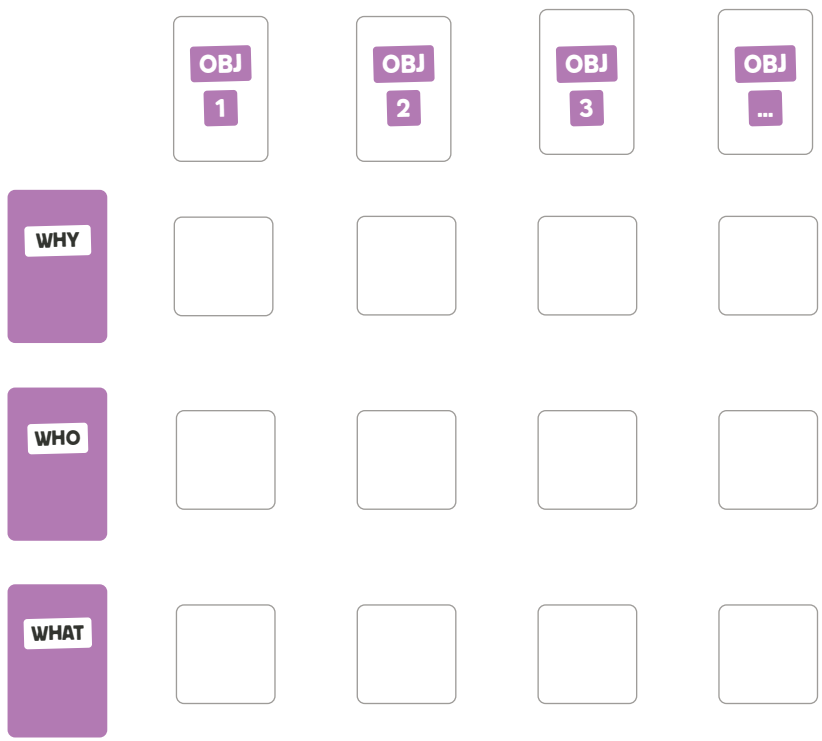
DEFINE OBJECTIVES

DEFINE OBJECTIVES

For each objective, address the questions: *Why*, *Who*, and *What*, where *Why* and *Who* will often shape *What* you do.

Create as many objectives as possible, ensuring every member's role and expertise is represented in at least some of them.

Write each answer on a post-it and place them in a grid like the one below.



STEP 1

OBJECTIVE

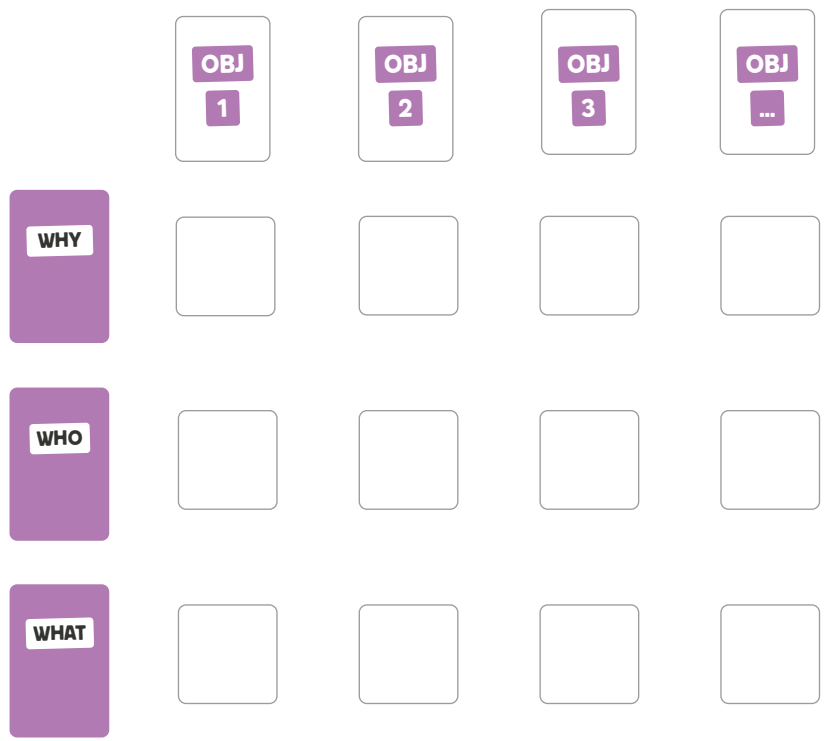
DEFINE OBJECTIVES

DEFINE OBJECTIVES

For each objective, address the questions: *Why*, *Who*, and *What*, where *Why* and *Who* will often shape *What* you do.

Create as many objectives as possible, ensuring every member's role and expertise is represented in at least some of them.

Write each answer on a post-it and place them in a grid like the one below.



STEP 1

WHAT

DEFINE OBJECTIVES

WHAT DOES THE RECOMMENDER DO?

What does your recommender system try to accomplish? Try to think beyond providing people with content they would like to click.

QUESTION

What do your users want from your recommender system beside seeing more of what they already like? What goals do other stakeholders have with the recommender system?

FOR EXAMPLE

- For a **news recommender:** providing an impartial overview of the news.
- For a **music recommender:** allowing new creators to find their audiences.
- For a **social media platform:** giving a platform to underrepresented groups.
- For a **government:** informing citizens of what is going on in their neighborhood.

STEP 1

WHY

DEFINE OBJECTIVES

WHY DOES YOUR ORGANIZATION WANT TO RECOMMEND CONTENT?

Many organizations have a mission, or core values driving them. In parallel, business goals need to be met, and users need to be happy with the system.

QUESTION

Why does your recommender system aim to accomplish this specific objective?

FOR EXAMPLE

BBC → To act in the public interest, serving all audiences through the provision of impartial, high-quality and distinctive output and services which inform, educate and entertain.

Netflix → To entertain the world.

STEP 1

WHO

DEFINE OBJECTIVES

WHO DOES THE RECOMMENDER RECOMMEND FOR?

Besides the users of the system, there are many parties (stakeholders) that could be considered. These people may have different requirements at different parts of the day.

QUESTION

Who do you think of when fulfilling this objective?

FOR EXAMPLE

- **People** on their way to work.
- **Creators** trying to find new audiences.
- **Advertisers** trying to reach their loyal customers.
- The **newsroom** upholding their journalistic and editorial values.

STEP 2

PRIORITIZE

PRIORITIZE

Which of the objectives defined in **Step 1** should be prioritized and worked on first? Ensure that your role/discipline is represented, but also consider what is most important to your organization as a whole, and what is important to people that are not currently at the table.

- Each group member gets **5 votes**.
- Distribute your votes among the objectives you find most important by tallying votes on post-its.
- One person can vote for an objective multiple times.

EXAMPLE

- The **newsroom** prioritizes objectives focusing on a balanced news offer.
- The **marketing team** prioritizes objectives that make the recommender system more likely to generate additional income.

STEP 3

DESIGN METRICS

BUILD EVALUATION METRICS

Build evaluation metrics that reflect the objectives you defined in the previous step. Go through the objectives you wrote one by one, starting with the ones that were voted highest.

A metric consists of three parts:

1. *What* needs to be measured.
2. *Where* you want to measure.
3. The *value* you expect the measurement to take.

For each objective, combine appropriate cards. Use the provided **metric builder** during your discussions. When you are satisfied with a metric, give it a name and write the name and titles of the chosen cards down.

→ **Examples**: see the information sheet for an example.

STEP 3

CATEGORY
GENRE
TOPIC

ITEM

CATEGORY, GENRE, TOPIC

A categorization of the content on the item being recommended. Diversification may help to avoid saturation and to keep a user engaged, or to give them a good overview of all the content that exists in the system.

EXAMPLE

- A **news organization** may want to ensure that readers have seen all the important topics from that day.
- A **travel agency's** recommender system may want to mix hotels, hostels and appartments.

STEP 3

STYLE

ITEM

STYLE

How a message is communicated and how easily users can access it. Relevant factors include but are not limited to content complexity, the time users must (and are willing to) invest, the language used, the target audience, or the medium (text, video, or audio).

EXAMPLE

- Ensuring that somebody that struggles with reading long texts gets recommended **shorter items** rather than lengthy ones.
- Recommending short-form **videos** to younger audiences.

STEP 3

PUBLICATION DATE

ITEM

PUBLICATION DATE

The date and time when the item was published. Some recommendations may prioritize newer content, whereas for others this would not matter as much.

EXAMPLE

- A recommendation on a **news organization's front page** may never recommend articles from more than 6 hours ago, whereas a recommendation showing events similar to the one the user is currently reading about would.
- A **library** recommending books from a certain theme, ensuring that books from different time periods are included.

STEP 3

QUALITY

ITEM

QUALITY

The quality of an item as determined by the organization that generates the recommendation. A recommendation could prioritize high-quality or important content over a users' preferred content types.

EXAMPLE

- A **library** may also want to recommend serious literature even when a user only reads comic books and manga.
- A **news organization** may prioritize high quality content on top of the recommendation, and allow for more entertaining content further down.

STEP 3

POPULARITY

ITEM

POPULARITY

How many other users have engaged with this item, or are expected to. In some cases popular items are likely to be items that readers are looking for, in other cases people may have already seen them.

EXAMPLE

- Recommending gossip news may result in a lot of clicks in the short term, but lead to user disappointment and loss **in the future**.
- Helping viewers find the shows they most likely entered the system for to find, which often are the **more popular** shows.

STEP 3

SUBJECT

SUBJECT

Who is the subject of the content in the recommendation, either through active participation or by being discussed by others.

EXAMPLE

Monitor that the guests on a talkshow are **representative of the population**, or that they highlight a **specific demographic group** that is affected by current events in the news.

CONNECT TO A HUMAN CARD

STEP 3

CREATOR

CREATOR

Who created the items in the recommendations. Could be considered to ensure for example fairness of representation and recognizability to the users.

EXAMPLE

Ensuring that **content produced** by different demographic groups are included in the recommendation, or that the user receiving the recommendation can **recognize themselves**.

CONNECT TO A HUMAN CARD

STEP 3

USER

USER

The recommendation could consider to whom it is presenting the recommendation, and their personal characteristics. Sometimes this may mean reflecting the user, and sometimes it may mean doing exactly the opposite.

EXAMPLE

Expose users to **different** opinions than their own, or make users **recognize themselves** in the recommendation.

CONNECT TO A HUMAN CARD

STEP 3

VIEWPOINT

HUMAN

VIEWPOINT

A viewpoint is a particular perspective, stance, or position on an issue that is expressed in the content being recommended. A recommender system may want to balance viewpoints in the recommendation, as to not give more visibility to one over the other. Alternatively, the recommender may want to allow a user to deeply inform themselves on one specific perspective.

EXAMPLE

Visibility of different **political parties** or **ideological expressions**.

STEP 3

GENDER

HUMAN

GENDER

May refer to gender identity, gender expression, or biological sex. Often expressed as a male/female binary, though gender exists on a spectrum. This spectrum includes but is not limited to non-binary, gender fluid, and transgender.

EXAMPLE

Ensuring that **underrepresented groups** are well-represented in the recommendation.

STEP 3

AGE

HUMAN

AGE

According to the dictionary, ‘the period of time someone has been alive or something has existed.’

EXAMPLE

People from **different age groups** may want to interact with a recommendation differently; alternatively, ensuring not only middle-aged people are reported on in the news.

STEP 3

BACKGROUND

HUMAN

BACKGROUND

May encompass things such as cultural background, nationality, ethnicity or skin color. These terms are distinctly different but often used interchangeably, and warrant additional discussion on what is meant exactly.

EXAMPLE

General **representation** of people of color, **countering** existing **biases** in society, or representing **the user** receiving the recommendation.

STEP 3

SEXUAL ORIENTATION

HUMAN

SEXUAL ORIENTATION

“An inherent or immutable enduring emotional, romantic or sexual attraction to other people. Note: an individual’s sexual orientation is independent of their gender identity.”

(Source: Human Rights Campaign)

EXAMPLE

Representation of **LGBTQIA+** people.

STEP 3

GEOGRAPHIC LOCATION

HUMAN

GEOGRAPHIC LOCATION

Where a person is located in the world. May refer to where they are currently based, or what they consider ‘home’.

EXAMPLE

Coverage of **local news**, including sufficient reporting from **rural versus urban** areas.

STEP 3

ABILITY

HUMAN

ABILITY

A person’s physical or mental ability that influence the way in which they can participate in society.

EXAMPLE

Recommending **shorter** texts to those with issues in reading, or providing **in-depth content** on subjects that the users are an expert on.

STEP 3

OTHER

HUMAN

OTHER

Other characteristics that may express a membership of a human to a social group.

EXAMPLE

Religion, education, occupation, socio-economic status, class, ...

STEP 3

CURRENT EVENTS

WORLD

CURRENT EVENTS

What goes on in the world determines what is relevant for the user to see and what a user needs to consume to fulfill their information needs.

EXAMPLE

- An item giving **additional context** on a topic relevant to the current election cycle.
- Ensuring that a recommendation covers **all the important** news of the day.

STEP 3

SOCIETY

WORLD

SOCIETY

The world as it currently is, including its biases and existing power balances. A recommendation could aim to reflect society, for example by proportionally covering different societal groups, or to counter known issues, for example by accounting for bias that we know is present in data gathered.

EXAMPLE

- Giving a platform to **underrepresented voices**.
- Ensuring that the recommender system is **fair** towards all demographic groups.

STEP 3

ONE SINGLE RECOMMENDATION

LEVEL

ONE SINGLE RECOMMENDATION

It only matters what is included in one single recommendation. This is useful if you want to make sure that every single recommendation is of sufficient quality.

You may want to count items that are recommended first more strongly than those that are recommended more at the bottom; discuss this as well.

EXAMPLE

- Ensuring that the recommendation does not contain items that are **too similar** to each other.
- Ensuring that every recommendation contains **at least two** 'serious' items.

STEP 3

ALL RECOMMENDATIONS TO A USER

LEVEL

ALL RECOMMENDATIONS ISSUED TO A USER

Ensure that what a user has seen over time is of sufficient quality. It is okay if not every single recommendation is completely balanced, but we do want to make sure that what is shown over time meets our standards.

EXAMPLE

- Over time, the user should have seen **equal amounts** of content reporting on the different political parties.
- The user should have seen **a good mix** of popular and niche content.

STEP 3

SYSTEM PERFORMANCE

LEVEL

SYSTEM PERFORMANCE

Choose this card when you want to monitor for effects at the system-level, rather than only within the recommendations generated.

Note that it will be difficult to determine whether the changes can be attributed to the recommender system, or that they are caused by something else.

EXAMPLE

- **Increase the number** of young users on the platform.
- **Increase the visibility** of niche and unpopular items.

STEP 3

USER CHARACTERISTICS

CONTEXT

USER CHARACTERISTICS

The degree in which a recommendation does or does not differ from who a user is or what they have consumed in the past.

EXAMPLE

Matching the users preferences, or on the contrary, presenting them with content types they **have not seen before**.

STEP 3

RECOMMENDATIONS TO OTHER USERS

CONTEXT

RECOMMENDATIONS TO OTHER USERS

The degree in which a recommendation matches with what other users have seen.

EXAMPLE

Preventing filter bubbles by ensuring that there are certain items **all users** have seen.

STEP 3

ORGANIZATION

ORGANIZATION

The degree in which the recommendation matches the organization's norms and values, including for example the general breadth of content the organization produces.

EXAMPLE

Ensuring that the recommendation contains **all item categories** the organization publishes.

CONTEXT

STEP 3

OUTSIDE WORLD

CONTEXT

OUTSIDE WORLD

The degree in which the recommendation matches the world in which it operates.

EXAMPLE

Political representation reflective of the size of the parties currently in government, or whether the recommendation covers **all the relevant news** of the day.

STEP 3

UNIFORM

UNIFORM

The degree in which a recommendation matches a uniform distribution; all aspects are present equally.

EXAMPLE

All political parties need to get **exactly the same amount** of visibility, or all item categories need to be in the recommendation **the same number of times**.

CONTEXT

STEP 3

SIMILAR

IN RELATION TO

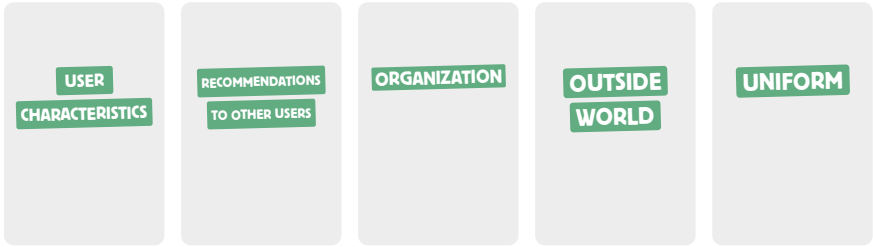
↓

CONNECT TO A CONTEXT CARD

SIMILAR

We expect the recommendation and the context to follow a **similar distribution**.

Choose the relevant **context card** (these are white with green), and place them on the white rectangle on the front side of this card.



EXAMPLE

The distribution of topics in the recommendation is **very similar** to the distribution of topics in the user's preferences, in order to help them find the content that you know they will be interested in.

STEP 3

DIFFERENT

IN RELATION TO

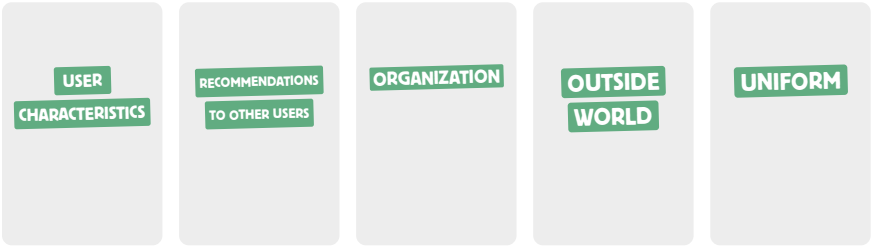
↓

CONNECT TO A CONTEXT CARD

SIMILAR

We expect the recommendation and the context to follow a **different distribution**.

Choose the relevant **context card** (these are white with green), and place them on the white rectangle on the front side of this card.



EXAMPLE

The distribution of topics in the recommendation is **very different** to the distribution of topics in a user's preferences, in order to help them find new interesting content they were not aware of before.

STEP 3

%

VALUE

PERCENTAGE

A certain percentage of the recommendation needs to be of a particular type.

EXAMPLE

We need **at least half** of the recommendation to consist of ‘hard news’ items.

STEP 3

HIGHER IS BETTER

VALUE

HIGHER IS BETTER

The more a particular aspect is present in the recommendation, the better.

EXAMPLE

The more niche items are recommended, **the better**.

STEP 3

LOWER IS BETTER

VALUE

LOWER IS BETTER

The less a particular aspect is present in the recommendation, the better

EXAMPLE

The less items produced by other organizations are present in the recommendation, **the better.**

STEP 3

NUMERIC

VALUE

NUMERIC

There needs to be an exact number of a particular aspect in the recommendation.

EXAMPLE

The recommendation needs to include **one paid item**.

STEP 4

NOW, HOW, WOW

NOW, HOW, WOW

Use the **now/how/wow matrix** to sort the designed metrics by:

1. How easy they are to implement.
2. How potentially impactful they are.

NOW Easy and clear to implement, data is already available, could be done relatively quickly.

HOW Good ideas but not ready yet; need more research, tools, money, or skills.

WOW Great ideas with a lot of value that should be picked up as soon as possible.

LATER/DREAM If something feels almost impossible with today's resources, park it in a 'Later/Dream' corner.

STEP 5

NEXT STEPS

NEXT STEPS

Look at the **now/how/wow matrix** you created. Who is responsible for taking the next steps towards implementation?

Write down the name of the objective, the name of the person or team responsible, and the specific action they will take.

EXAMPLE

Data science needs to research whether items can be clustered into topics. **Legal** needs to investigate whether under GDPR data point X can be used. **Editorial** needs to find out which categories should be prioritized.

Or, all data and information is already present, and the **recommender system developers** need to build the metric into an existing dashboard.

STEP 6

WRAP UP

WRAP UP

Reflect on the process as a whole.

- What went well?
- What would you do differently if you were to start again?
- Has everyone's perspective been sufficiently heard?
- Were any perspectives structurally missed?
- When will the actions identified in **Next Steps** be completed?

Take these insights with you as you continue working on your recommendations.

VALUES TO METRICS TOOLKIT

AI, MEDIA & DEMOCRACY LAB
UNIVERSITEIT VAN AMSTERDAM

VALUES TO METRICS

A card-based workshop tool for news organizations to reflect on and improve their recommendation systems. The steps guide participants from defining objectives to designing metrics and identifying next steps.

CONCEPT & RESEARCH

Developed within the AI, Media & Democracy Lab. Idea and primary research by **Sanne Vrijenhoek** and **Sara Spaargaren**.

CONTRIBUTORS

Kornelija Gruodyte, Lien Michiels, Savvina Daniil, Pascal Wiggers, Maaike Harbers, Natali Helberger.

VISUAL CONCEPT & GRAPHIC DESIGN

Daphne de Vries / Bureau Merkwaardig

This toolkit is part of the AI, Media & Democracy Lab. For more information about the lab and its further activities, visit **aim4dem.nl**