Authorship, plagiarism and entitlements to data in the age of AI
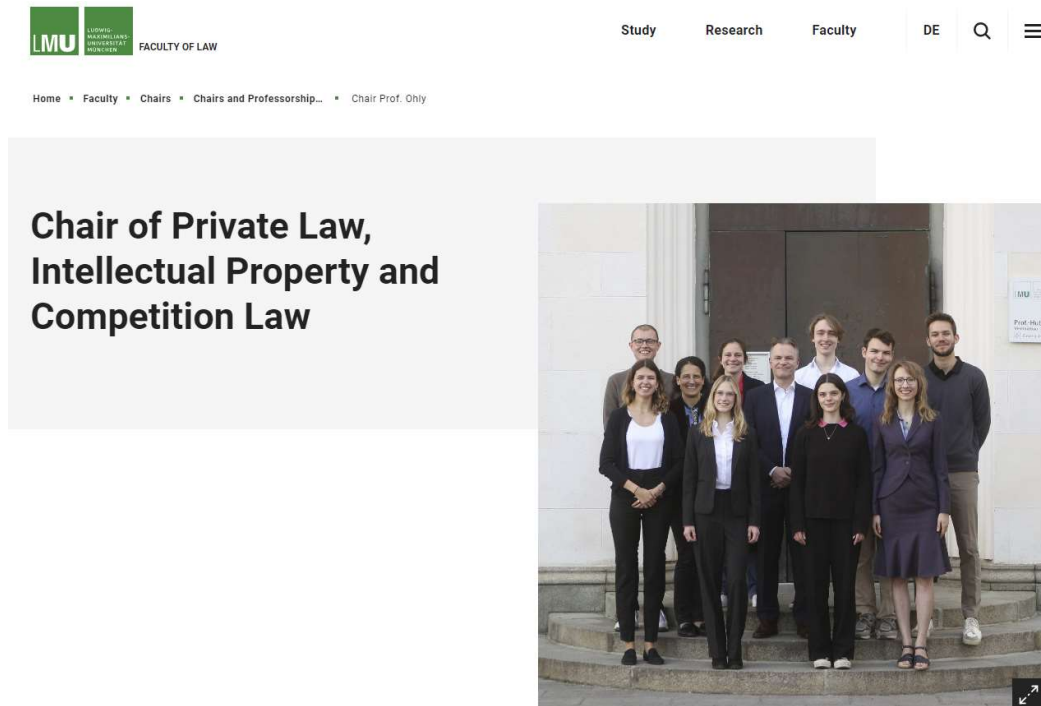
Prof. Dr. Ansgar Ohly, LL.M. (Cambridge)

Responsible Research 2024

Biocenter LMU, 25 July 2024

Why I am interested

Chair of Private Law, Intellectual Property and Competition Law

«Summa cum laude»
– *Universität Bayreuth*

And why you should be interested

§ 1(2) LMU Ordinance for the Safeguarding of Good Academic Practice

"All those working in science and research at LMU are obliged and responsible for complying with the rules of good academic practice in their conduct."

Agenda

Authorship → Plagiarism → Entitlements to data → In the age of AI
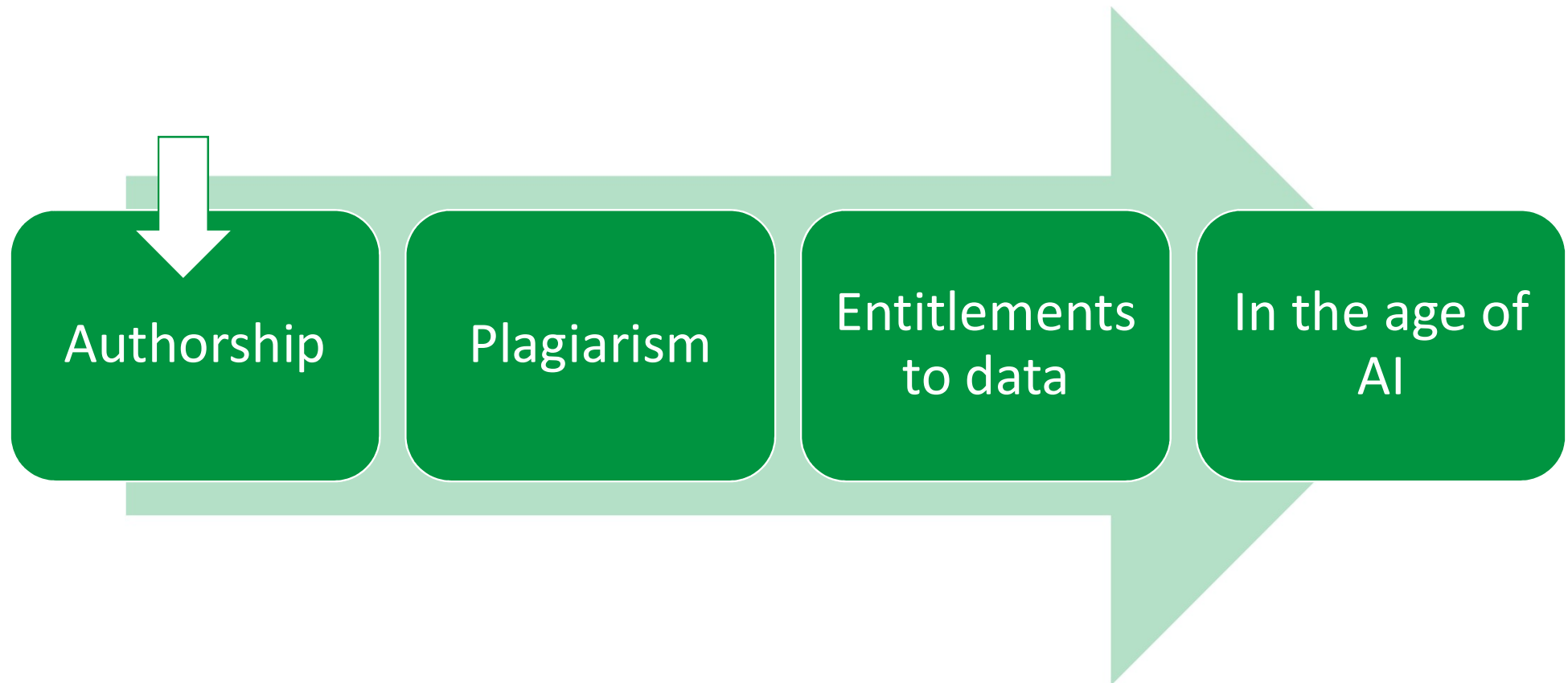
1. Authorship

Authorship and plagiarism

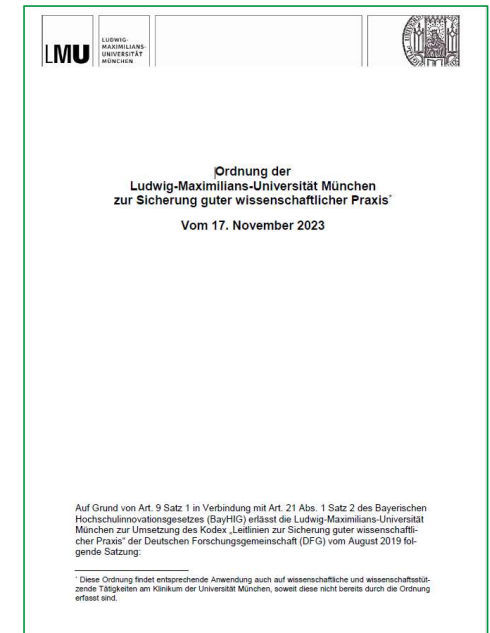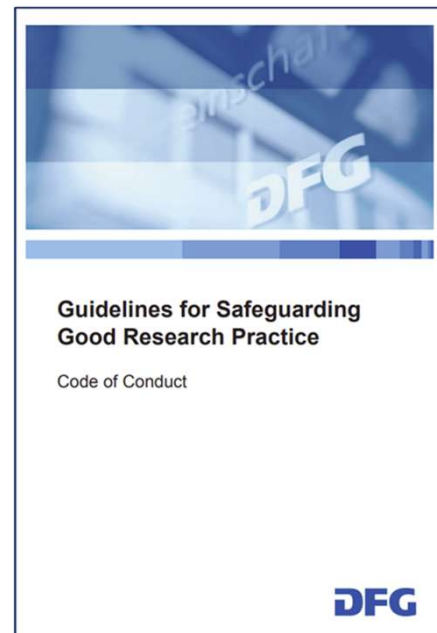**Good academic practice**

Copyright law

1. Authorship

Authorship

Good Research Practice

- Self-regulation
- Guideline 14 DFG Code: 'genuine, identifiable contribution to the content of a research publication'
- Fair attribution of *research results*
- Violation → self-control procedure of university or research institution

Guidelines for Safeguarding Good Research Practice

Code of Conduct

DFG

Ordnung der
Ludwig-Maximilians-Universität München
zur Sicherung guter wissenschaftlicher Praxis*
Vom 17. November 2023

Auf Grund von Art. 9 Satz 1 in Verbindung mit Art. 21 Abs. 1 Satz 2 des Bayerischen Hochschulinnovationsgesetzes (BayHIG) erlässt die Ludwig-Maximilians-Universität München zur Umsetzung des Kodex „Leitlinien zur Sicherung guter wissenschaftlicher Praxis" der Deutschen Forschungsgemeinschaft (DFG) vom August 2019 folgende Satzung:

* Diese Ordnung findet entsprechende Anwendung auch auf wissenschaftliche und wissenschaftsstützende Tätigkeiten am Klinikum der Universität München, soweit diese nicht bereits durch die Ordnung erfasst sind.

1. Authorship

Authorship

Good Research Practice

- Self-regulation
- Guideline 14 DFG Code: 'genuine, identifiable contribution to the content of a research publication'
- Fair attribution of research results
- Violation → self-control procedure of university or research institution

Copyright law

- (national) law
- § 7 UrhG: ,Author is the creator of the work'
- Work: expression (ie text), not idea
- Freedom of scientific content, attribution of economic and personal rights *in individual expression*
- Infringement → lawsuit

1. Authorship

Who is the author?

- Disputes over authorship are "frequent, harsh and ugly" (*Robert K. Merton*)
- An author is an individual who has made a genuine, identifiable contribution to the content of a research publication of text, data or software (G 14).
- Co-authors
  - take joint responsibility
  - agree on final version
  - identify, as far as possible, exact nature of their contribution
- Order of authors depends on (1) agreement and (2) discipline

1. Authorship

Who is the author?

- You are a (co-)author when you
  - develop and design the research project
  - gather, collect, acquire data, software or sources [in more than a purely technical way]
  - analyse and interpret data, sources, etc
  - write the text
- You are not a (co-) author just because you
  - are the head of the research unit or the supervisor
  - have secured funding
  - have provided the standard methodology
  - Have read the manuscript and given purely editorial comments
  - make purely technical contributions

1. Authorship: case study

Post-docs A, B and C design a research project, conduct experiments, interpret the results and write a publication. Professor P, who is the head of the institute, only finds out about this when the text has already been finalized. He makes editorial comments on the text and suggests clarifications and corrections of minor errors.

1. P requests to be cited as the last author and threatens A, B and C with negative consequences if they do not agree.

2. A, B and C consider mentioning post-doc D, with whom they discussed the product at an early stage, because D is in need of another publication supporting her application for a permanent university post.

Agenda

Authorship → Plagiarism → Entitlements to data → In the age of AI
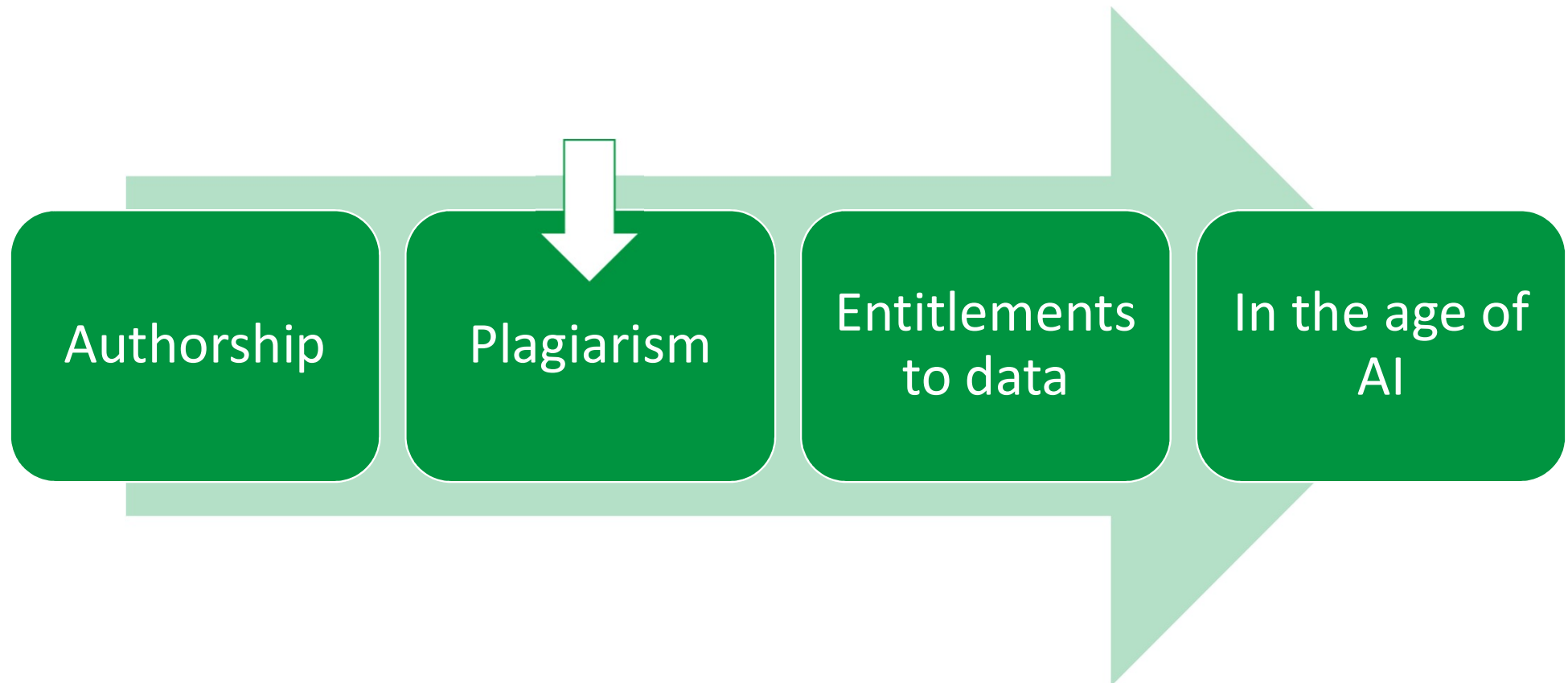
2. Plagiarism

Plagiarism defined

- Martial (40-104 A.D.): those who steal other people's verses are as bad as those who steal other people's slaves ("plagiarii")

- Passing off someone else's work as one's own = misappropriation of someone else's work, combined with the misrepresentation of one's own authorship

- But no legal definition of plagiarism

2. Plagiarism

Plagiarism = just one of several forms of academic misconduct

Manipulating experiments / faking results

Insisting on wrong order of authors

Passing of someone else's text / ideas as your own

Destruction / insufficient storage of experimental data

Misappropriating ideas in your role as a referee

Sabotaging other scientists' work

Plagiarism = a concept with a clear core and furry edges

"copy & paste" of text

Sloppy referencing

adoption of text in rephrased form

Selective citation for bad reasons

misappropriation of ideas

Selective citation for good reasons
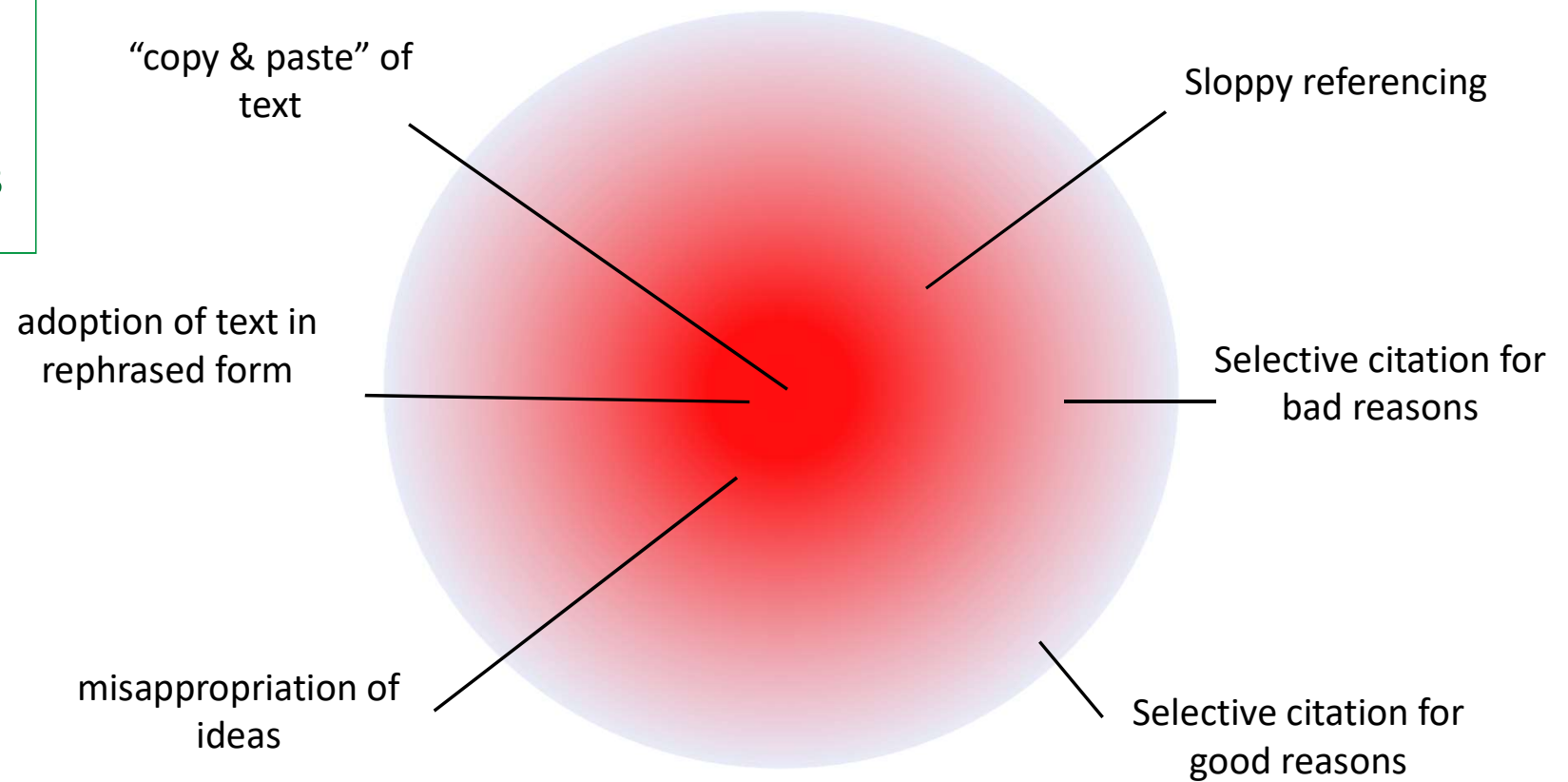
2. Plagiarism

Plagiarism defined

- No legal definition

- Misappropriation of someone else's work, combined with the misrepresentation of one's own authorship

- Clear core meaning, but furry edges

- Depends on …
  - citation practices in respective academic discipline
  - type of work: in-depth study or overview

- Qualitative evaluation v. quantitative results of automated processes

- How about self-plagiarism?

2. Plagiarism

Consequences

- Legal
  - If the text (as opposed to the mere results) has been taken → copyright infringement
  - Author must sue
  - Injunction + damages (criminal sanctions are rare)
- Good academic practice
 - Self-control mechanism of LMU or research institution

2. Procedure in case of misconduct

Step 1
- **Ombudsperson (G 6, 19)**
- Advises on good practice
- Takes complaints
- Conducts preliminary investigation

Step 2
- **Investigation Committee (G 19)**
- Decides whether to open proceedings
- Investigates in accordance with rules of law and fairness
- Makes recommendations to governing body

Step 3
- **Governing body (G 3, 19)**
- Administers sanctions on the basis of Investigation Committee's recommentations

Rules of procedure (R 18, 19)

- Procedure confidential
- Measures to protect both complainant and respondent
- Presumption of innocence
- Both have the right to be heard at all stages
- False allegations may constitute academic misconduct

Agenda

Authorship | Plagiarism | Entitlements to data | In the age of AI
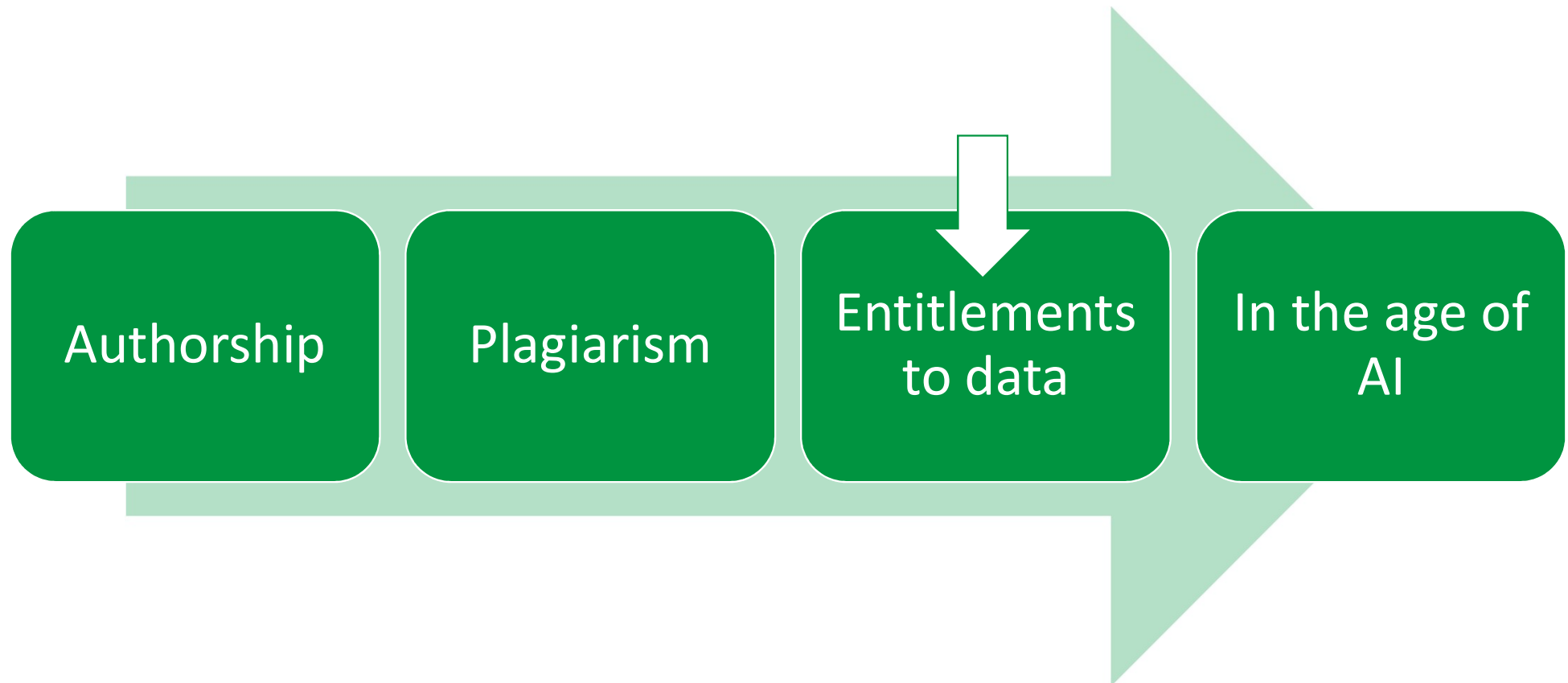
3. Entitlements to data

```
                        ┌─────────────┐
                        │    Data     │
                        └─────────────┘
                       /               \
```

Good Research Practice

- Documentation, archivating, making available

- No clear principles on 'ownership'

Law

- No IP right to data

- © protection of databases → protection against copying etc of database, not of individual data

- Datasets can qualify as trade secrets → use or disclosure can be infringements, but several exceptions and no clear definition of ownership

3. Entitlements to data

The DFG Code: dealing with data

- Origin of data must be disclosed (G 7)

- All information relevant to production of results must be documented (G 12)

- Research data and principal materials must be made available in recognised archives and repositories in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) (G 13).

- Data should be archived for usually 10 years (G 17)

3. Key issues: data

The DFG Code: entitlement and access

- "Who owns the data"? Neither the law nor the code provide a clear answer
- Law: © in texts and databases, trade secrets, but no IP right in data or datasets
- DFG acknowledges that much (or everything) depends on individual circumstances
- → documented agreements on usage rights recommended, particularly when non-academic partners or other academic institutions are involved (G 10)

3. Data: case study

P is a professor of psychology. R is a post-doctoral researcher at P's institute. On the suggestion of P, R conducts a survey among 1,000 university students on how they have been affected by the COVID pandemic. The study is financed by a DFG grant, for which P and R have jointly applied. The plan is for R to write an academic paper on the basis of the set of data.

R compiles the data, anonymises them and stores them both on the institute's computer and on her own laptop. But her research does not go well. She never finishes her paper. Her position at university ends, and R leaves P's institute after lengthy quarrels.

R thinks that the data belong to her because she has compiled them. She thinks that she is the only one who may use the data for her research, and that all papers based on these data must cite her as the co-author.
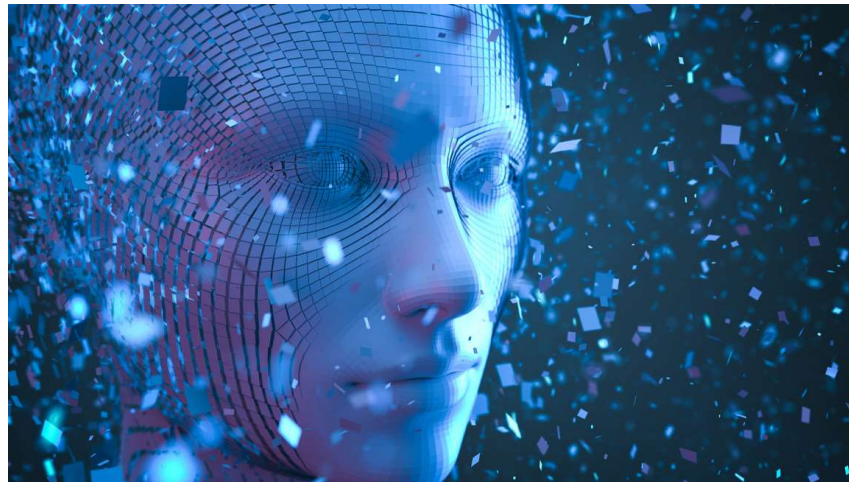
P plans another research project for which she would like to use the data. She thinks that she is entitled to do this and that there is no need to cite R as a co-author.

Who is right?

Agenda

Authorship → Plagiarism → Entitlements to data → In the age of AI
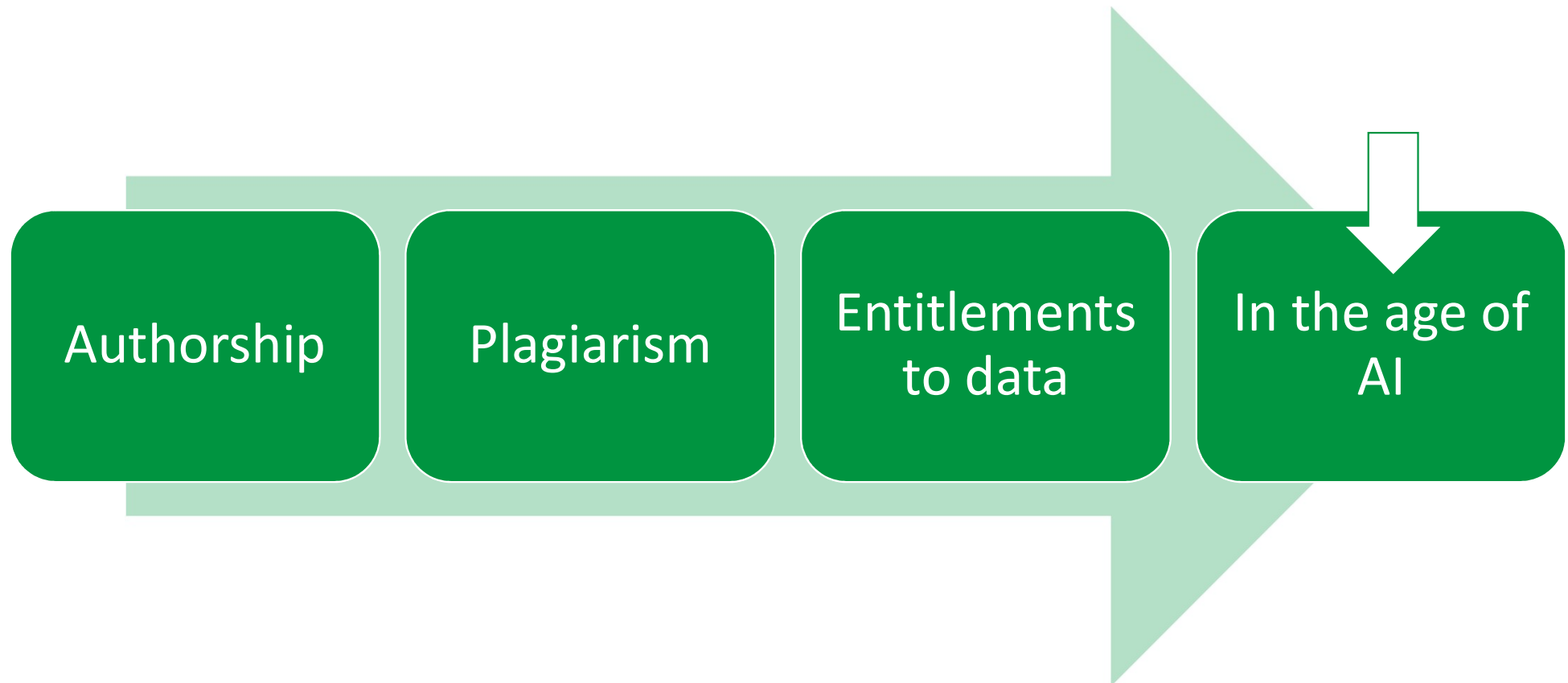
4. ... in the age of AI

???

tool   creator

4. … in the age of AI

The challenge: uncertain and potentially disappearing authorship

- Exponential development of AI systems

- Wide range of uses from tool to creation

- Widespread uncertainty

  - (To the best of my knowledge) no DFG statements

  - Universities and faculties have only started drafting guidelines

  - © law is of limited assistance because the first cases are only just emerging

- Does Professor Ohly know better?

- No, he doesn't.

- But there are some possible approaches.

Approach 1: copyright law

- © law requires a personal creation
- AI does not qualify as a (personal) creator
- But protection may arise if AI is only used as a tool
- The 1,000,000 $ question: Which quantity and quality of prompts is needed for the AI user to become an author?
- And by the way: infringement by training AI systems ('input') and by the AI output has become a big issue

Approach 2: draft LMU Law Faculty Guidelines

1. The use of AI tools for the design and writing of take-home exams and seminar papers is allowed. This does not release students from the general principles of academic work and the responsibility for the content and form of their work.

2. Students must refrain from entering their own or other people's personal data in AI tools.

3. For reasons of transparency, the use of AI tools should be referenced by the author in general form.

4. The use of AI tools as such does not have any influence on the assessment of the respective work.

5. The use and demonstration of AI tools in teaching, where appropriate, is encouraged.

4. … in the age of AI

Some tentative suggestions

- The use of AI systems as tools will become very common.

- You can still be an author if you merely use AI as a tool.

- The test is the same as in the analogue world: a genuine, identifiable contribution to the content of a research publication of text, data or software

  - asking original questions

  - interpreting data generated by AI

  - writing a text on the basis of an AI draft

  - But not just the use of simple / standard prompts

- But for the sake of transparency, the tools which have been used should be disclosed in appropriate (?!) detail.

Thank you very much for your attention!

1. Authorship

Authorship and publication

- Research results should be published (G 13)

  - Exceptions possible

  - Examples: patent application, personal data

- Mode and medium of publication depend on academic discipline (G 13, G 15)

  - Selection with due regard to quality and visibility in relevant field (G 15)

  - New media should be evaluated to assess their seriousness (G 15)

  - No clear commitment to "open access"

  - Software is made publicly available with the source code (G 13)