TARTU ÜLIKOOL

The extent of legal control over language data: the case of language technologies

The presenters:

Aleksei Kelli **Penny Labropoulou**

TARTU ÜLIKOOL
1632

The authors:

Aleksei Kelli, Arvi Tavast, Krister Lindén, Kadri Vider, Ramūnas Birštonas, Penny Labropoulou, Irene Kull, Gaabriel Tavits, Age Värv

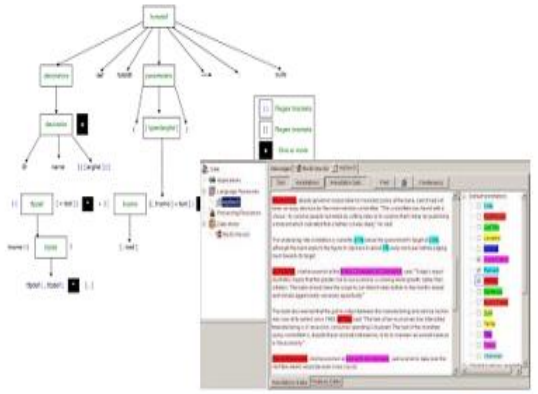
The aims of the research:

- To determine the extent of impact on the development of language technologies (LTs);
- To define a favourable regulatory framework for the development of LTs.

From Language Data to LTs:

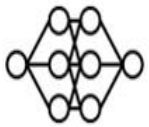
raw data

datasets



annotated data

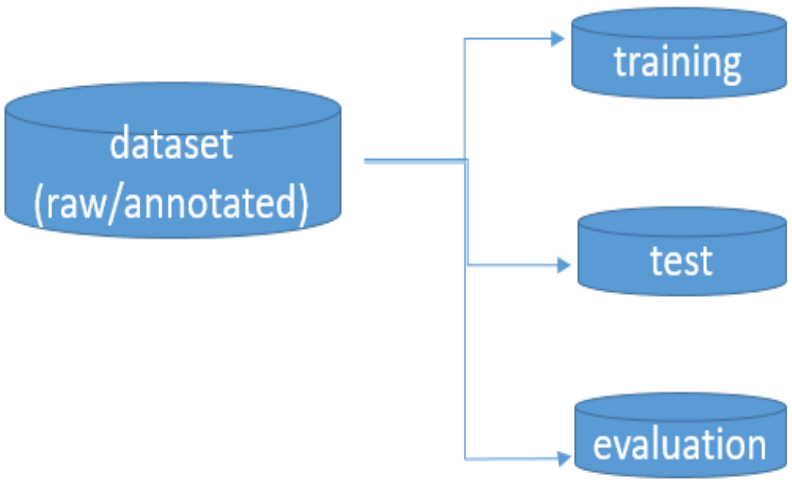
Reddit temporal n-gram corpus	word	frequency	year	month
1-gram	trump	981	2015	01
2-gram	trump apprentice	31	2015	01
3-gram	donald trump battle	16	2015	01
4-gram	donald trump ignorant tweet	8	2015	01
5-gram	take donald trump advice in	2	2015	01



"models"

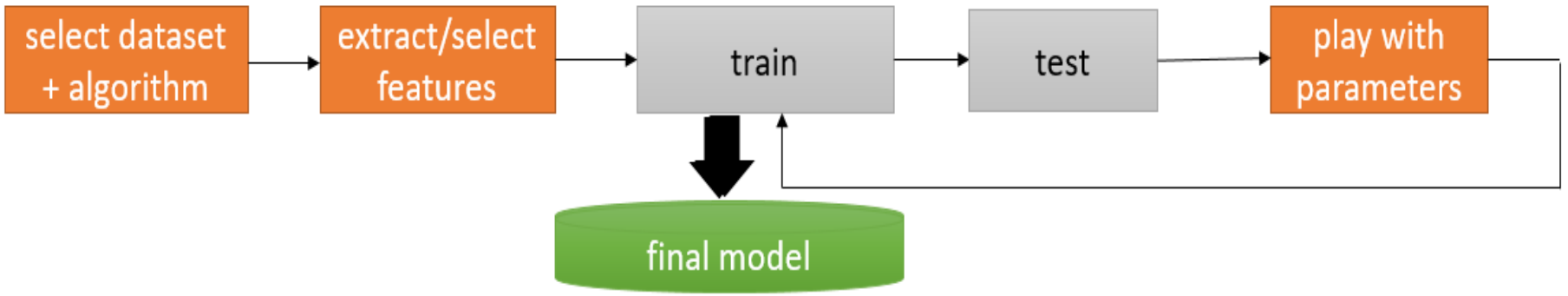
end-user applications

Building models process:

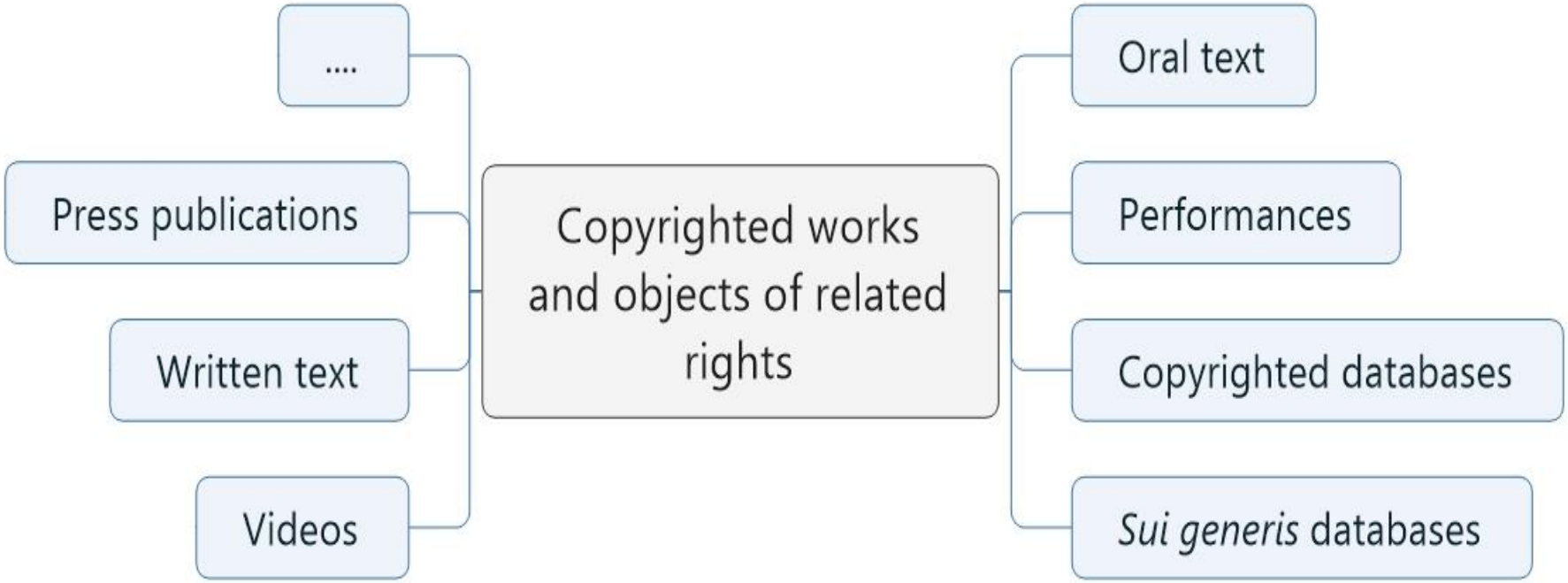


human-driven

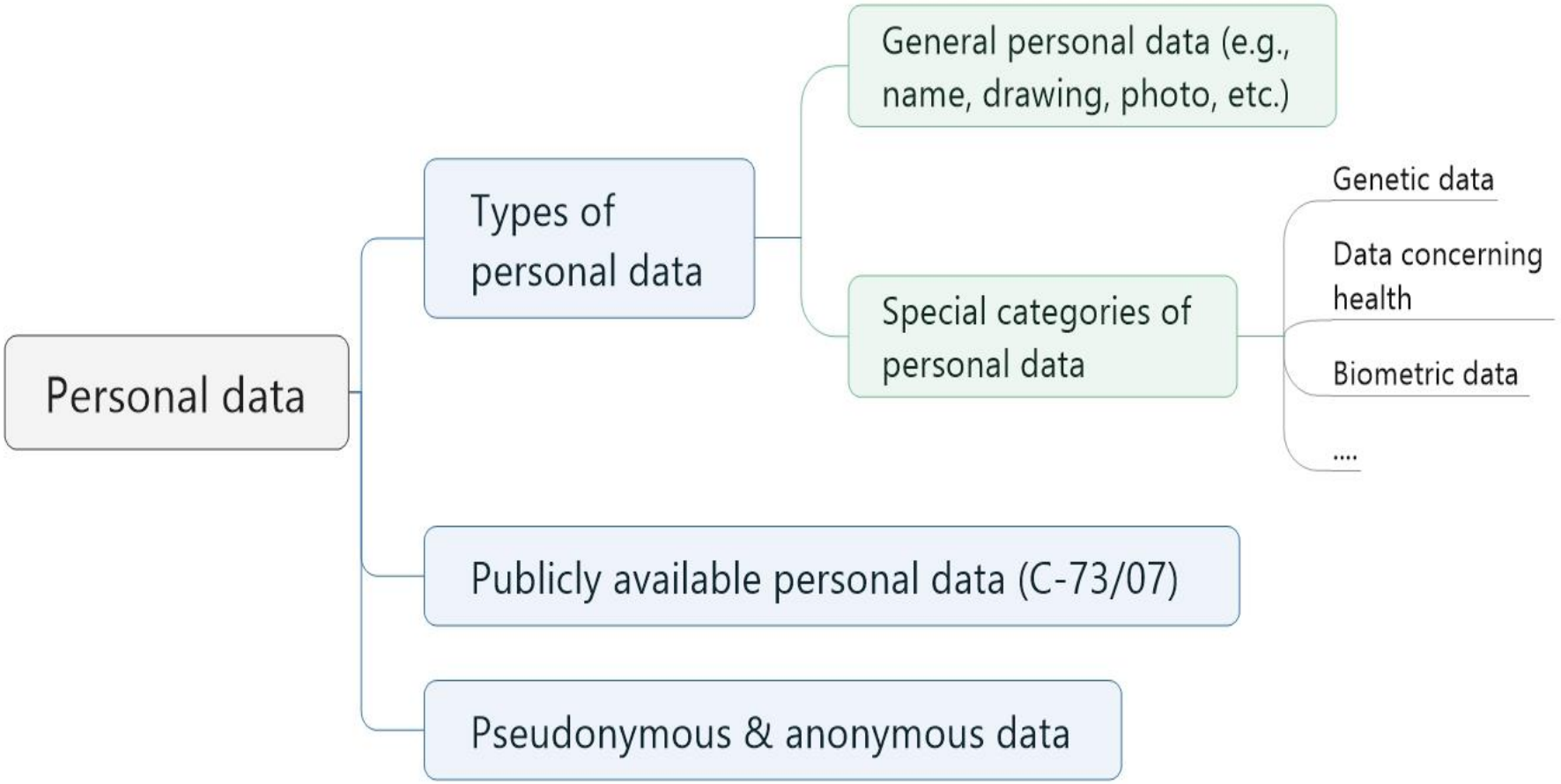
auto processing



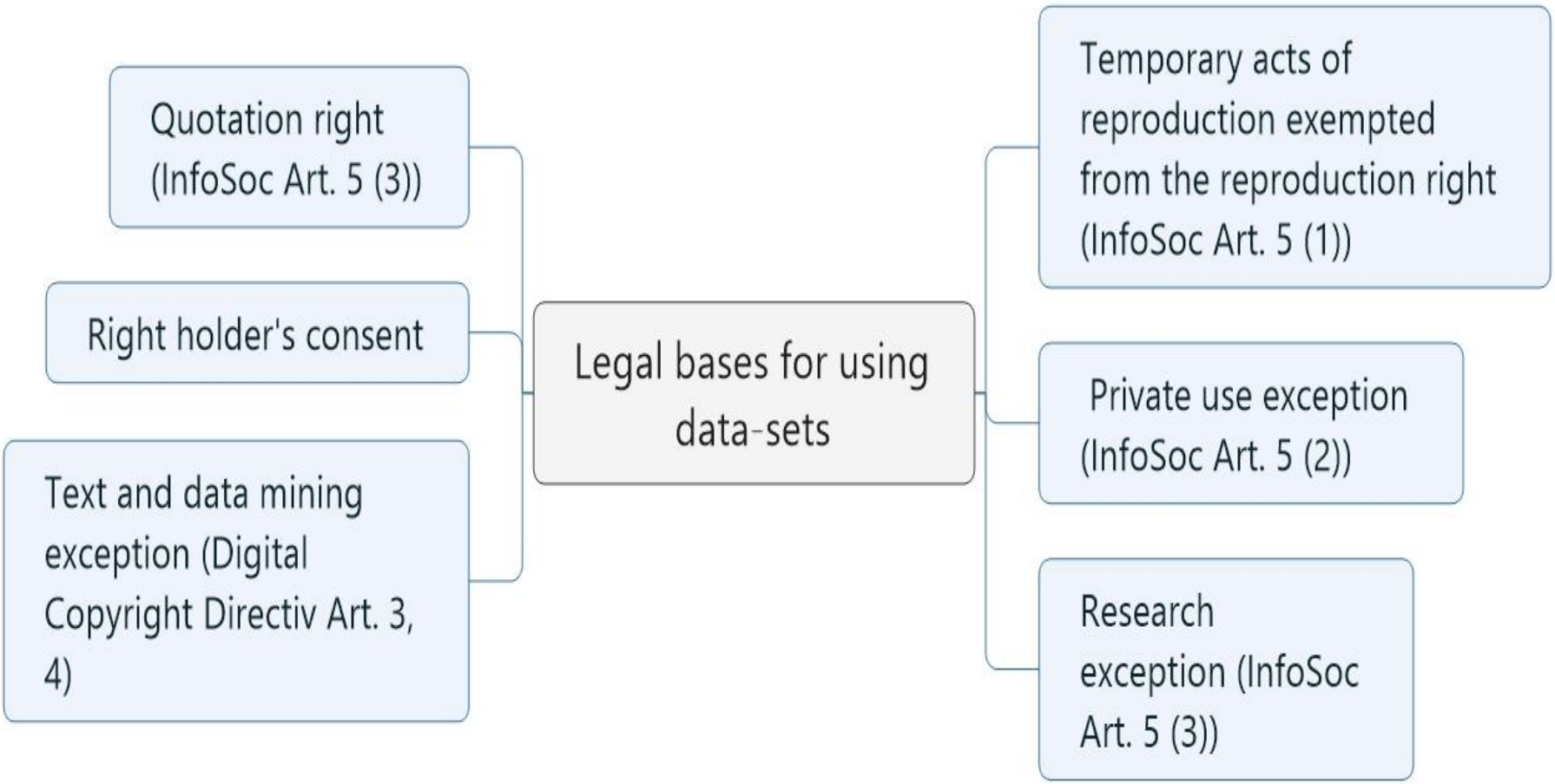
Language data & copyright:



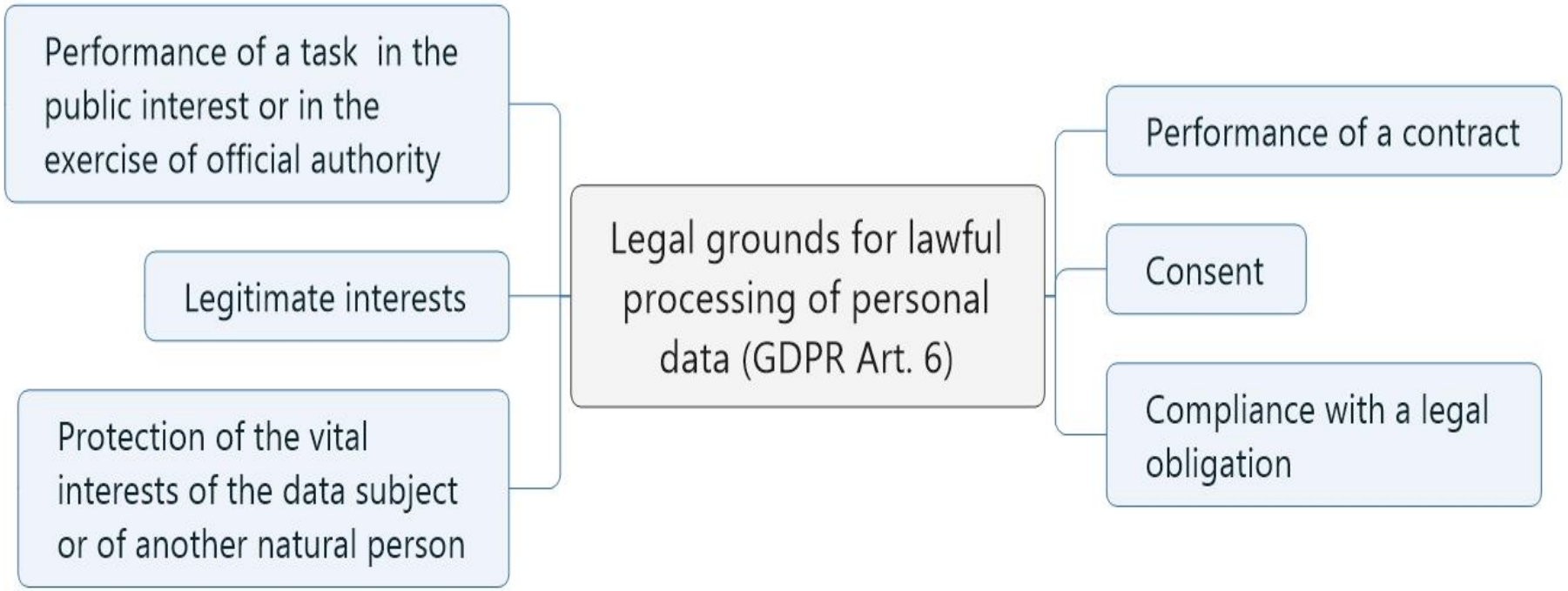
Language data & personal data:



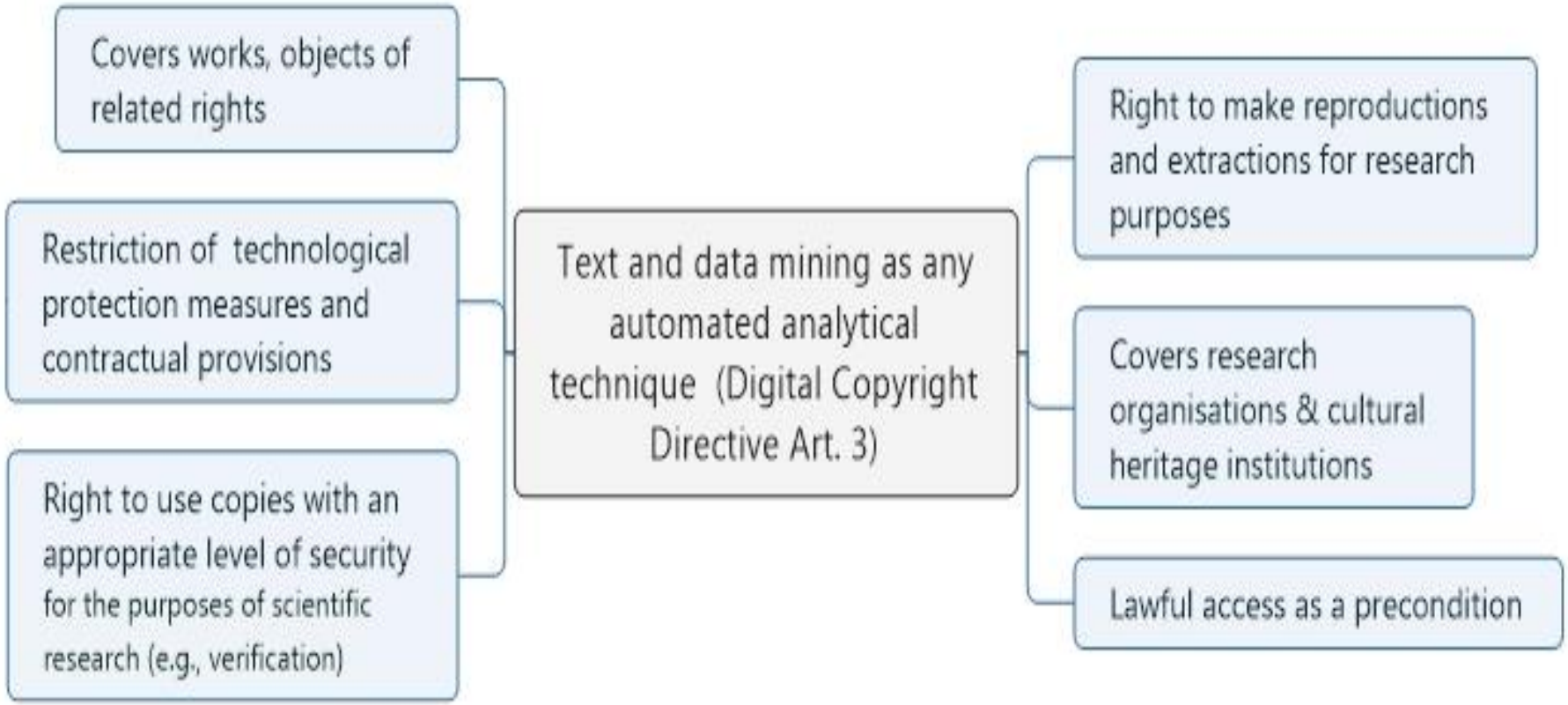
Use of LD & copyright protection:



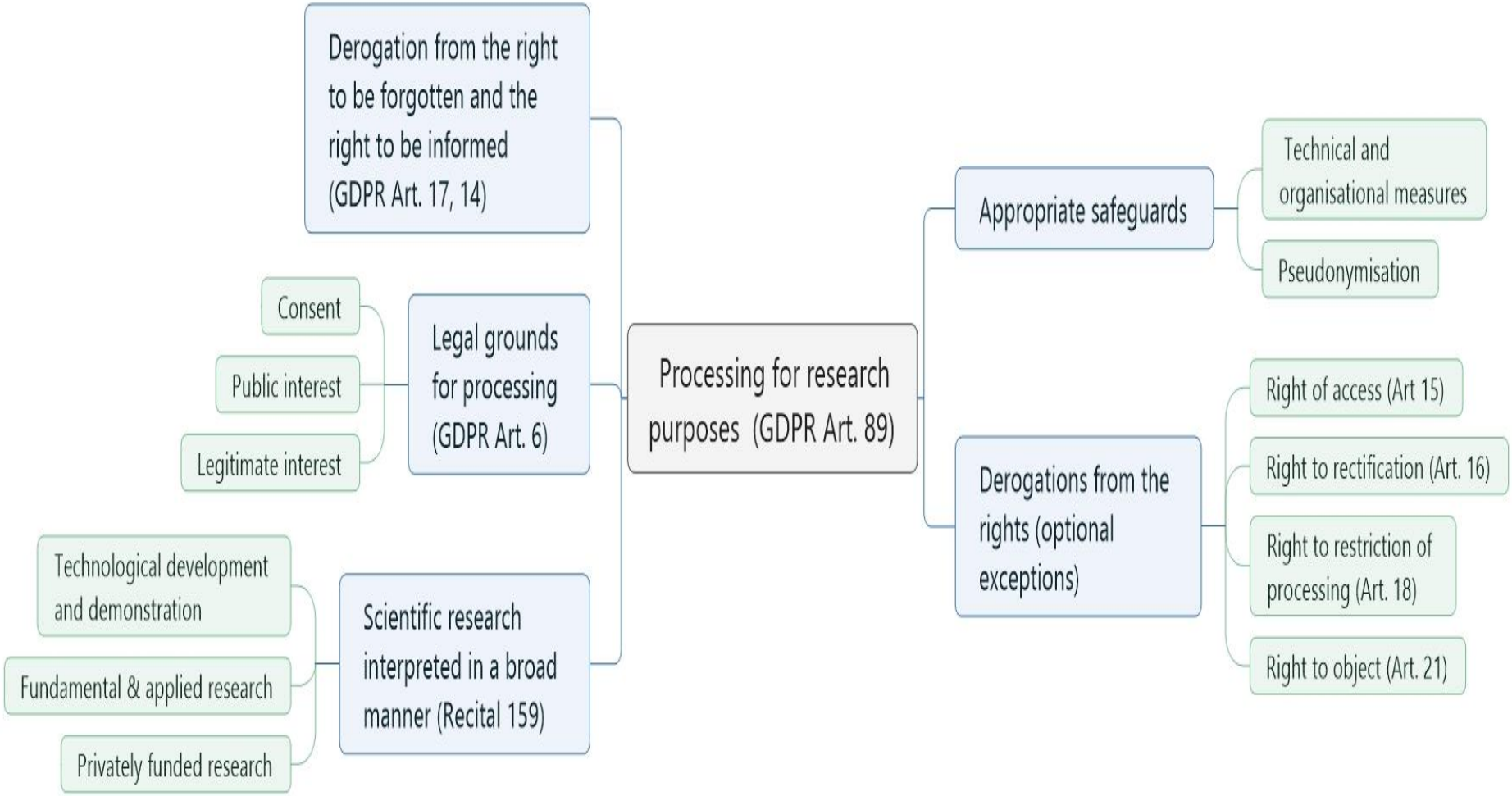
Use of LD & personal data protection:



Research use & copyright:

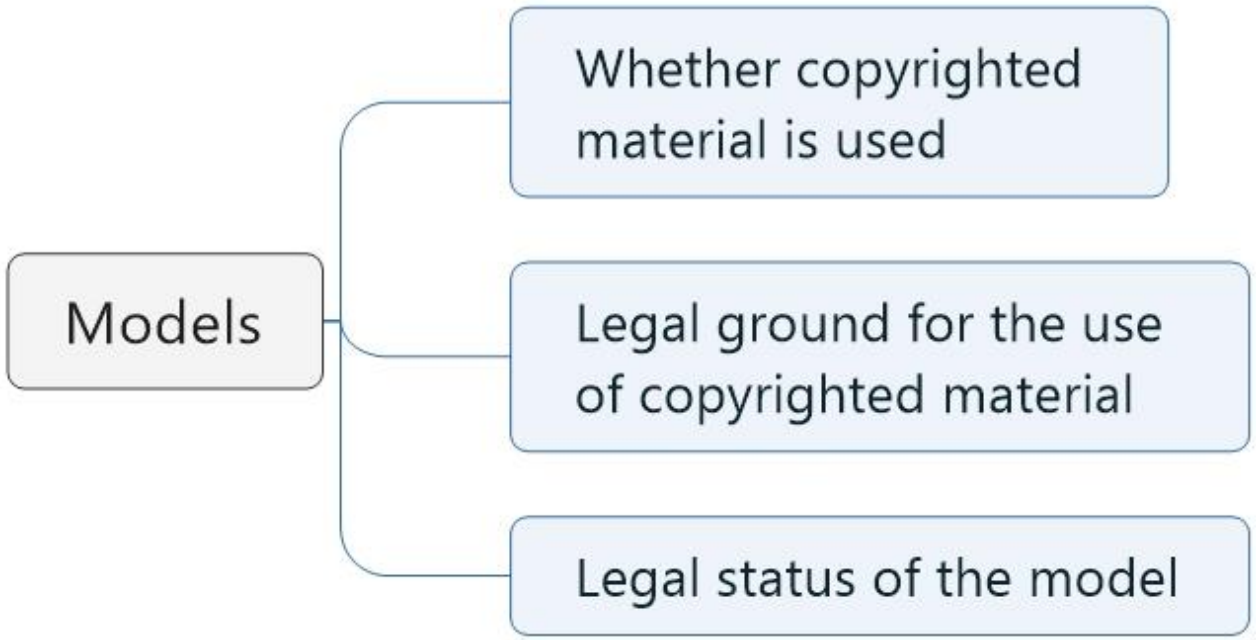


Research use & personal data:



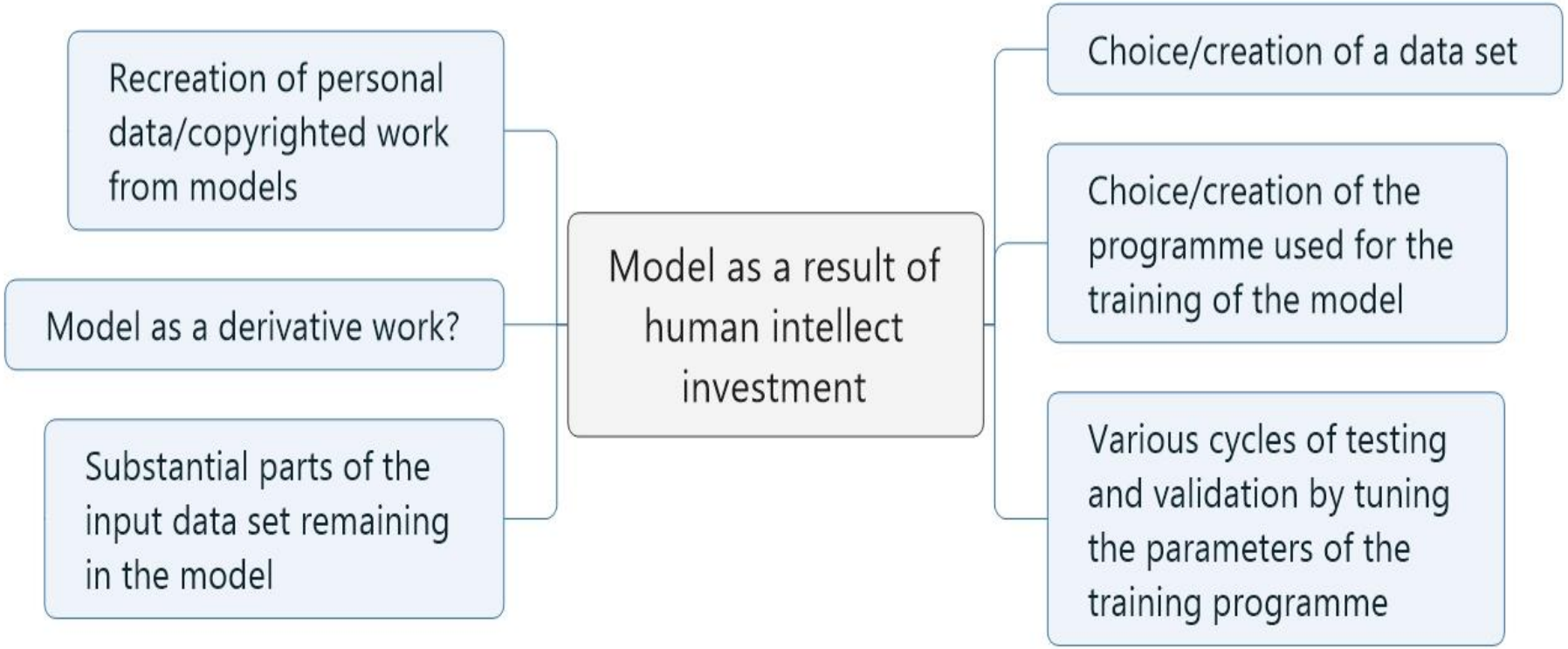
TARTU ÜLIKOOL

Models:

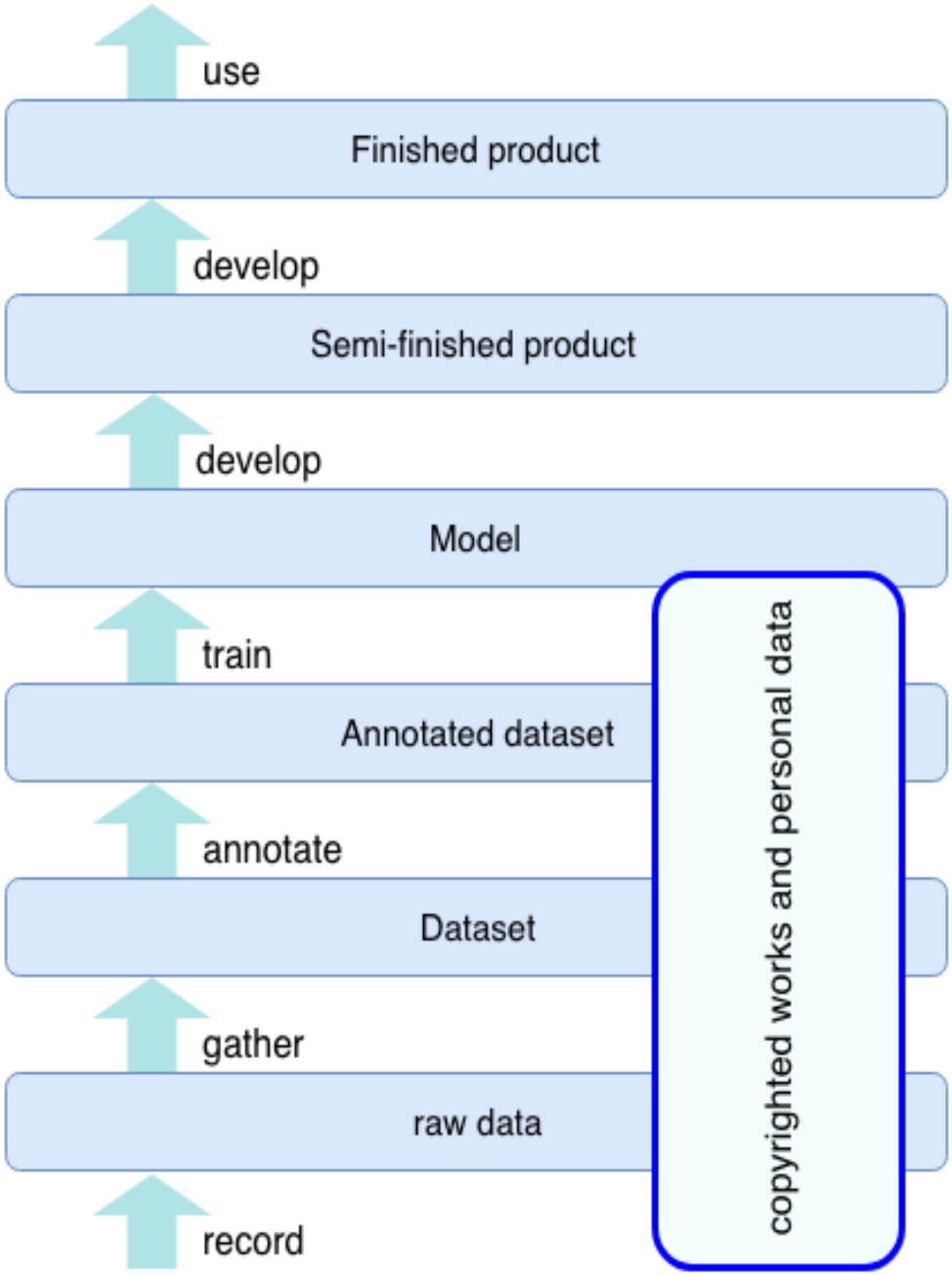


TARTU ÜLIKOOL

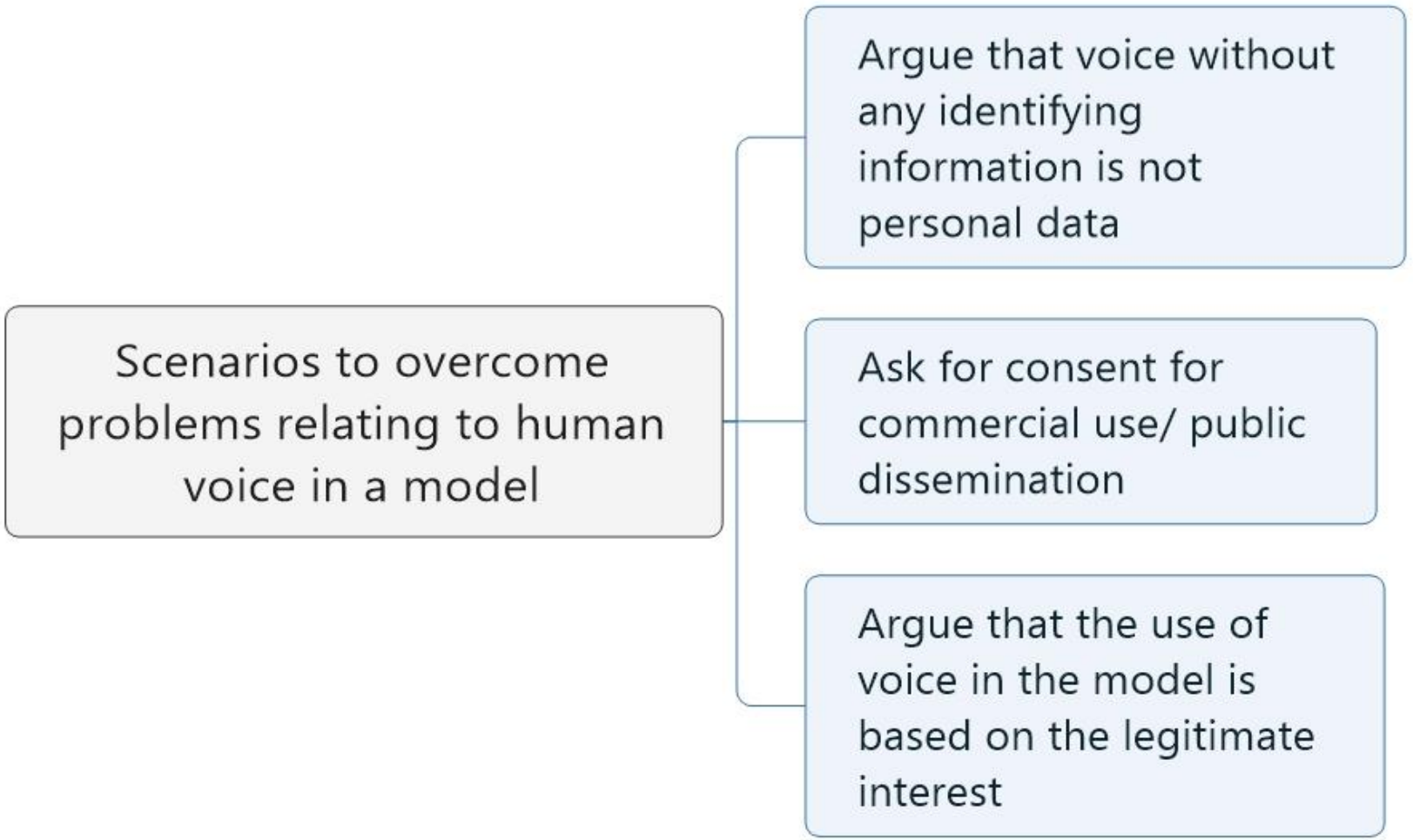
Models:



Models:


use
Finished product
develop
Semi-finished product
develop
Model
train
Annotated dataset
annotate
Dataset
gather
raw data
record
copyrighted works and personal data

Personal data problems concerning voice:



Conclusions:

- Models are not necessarily subject to the same copyright and personal data restrictions as language data used as input;
- The limited impact of copyright and personal data restrictions on models potentially allows to:
  - Rely on favourable copyright/personal data provisions supporting research;
  - Make models publicly available;
  - Commercial use of models;
  - Personal data problems with voice need to be addressed.

TARTU ÜLIKOOL
1632

Thank you