



**TILBURG SCHOOL OF HUMANITIES AND DIGITAL SCIENCES**  
**APPLICATION FORM: Ethics Review / Data Management / GDPR**

\*

**General guidelines of use**

- ❖ This submission form must be used for research projects which are going to be executed under supervision of TSHD affiliated researchers. Only principal investigators of a project can submit a project for evaluation. At least one of the principal investigators has a PhD and is either employed by Tilburg University or holds an endowed or honorary chair.
- ❖ The form consists of four parts. Part A is general information, Part B is the ethical review, Part C is the data management review, and Part D the General Data Protection Regulation (GDPR) review. Parts A and C are mandatory for all research projects within TSHD. Part D must be completed if the project deals with any personal data. The term 'personal data' should be interpreted as broadly as possible (see: <https://gdpr-info.eu/issues/personal-data/>). It is the responsibility of the principal investigator to decide if Part B needs to be completed. For more information (you must be logged in) <https://www.tilburguniversity.edu/intranet/organization-policy/erb/humanities>
- ❖ Ethical clearance of a research project is valid for the indicated duration of the project or until a change occurs in study population, data collection, or other procedures.
- ❖ The researchers and other involved personnel commit themselves to maximize the quality of the research, data analysis, and the reports and to respect specific rules and regulations concerning specific methodologies. The researchers and other involved personnel also guarantee that the study participant may discontinue their participation at all times without any consequences. Below mentioned researchers and other involved personnel commit themselves to treat all participants according to the most recent version of the [VSNU Netherlands Code of Conduct for Research Integrity](#). Moreover the researchers act in line with the [General Data Protection Regulation](#) (GDPR), and the [TiU Research Data Management Regulations](#).
- ❖ Please submit this form via email to: [tshd.redc@tilburguniversity.edu](mailto:tshd.redc@tilburguniversity.edu)
- ❖ With this (electronic) signature the undersigned declares to have described the research project truthfully.

For agreement:

Name (Principal Investigator): Emiel van Miltenburg

ANR (employee number):

Date: 26 April 2023



## Part A: General Information

**Title research project:** ReproHum (round 1) **Document Version:** 1

**Principal investigators:** Emiel van Miltenburg, Emiel Krahmer,

**Department:** DCC

**Co-applicant(s):** Frédéric Tomas, Nadine Braun, Anouck Braggaar, Saar Hommes, Chris van der Lee, Debby Damen, Martijn Goudbeek, Liesje van der Linden

**Department:** DCC

**Project duration:** 1 year until April 2023

**Funding organization:**

☐ The funding organizations are: [Click here to enter text.](#)

☒ There is no funding organization

### SUMMARY (Max. 1000 words)

Give a summary of the proposed research project. Ensure you give sufficient information on the data collection procedures (manipulations, stimuli, questionnaires, especially when they may be ethically sensitive), and the types of personal data you intend to collect.

#### 1.1 Background

This study is part of a general project called ReproHum (<https://reprohum.github.io/>), which aims to study the reproducibility of studies carried out in the field of Natural Language Processing. The Tilburg University Language Production (TULP) group joined this project as a partner lab, meaning that we carry out reproduction studies for published experiments. The goal here is to see how reliable and repeatable published results tend to be in our field. TULP has a long history of working on methodological issues in NLP, so this project is a natural extension of our prior research.

The first experiment that we aim to reproduce is from [Puduppully & Lapata \(2021\)](#). We will reproduce their experiment 2.

#### 1.2 Research question(s)

There are three questions that we aim to answer:

1. The meta-scientific question: can we produce the same results as the original study?
2. The original research question: which automatic summary generation system produces the best summaries?
3. Our own methodological question: what can we learn from studying the participants' responses to the items provided in the experiment? (E.g., how reliable are these ratings?)

#### 1.3 Study design and methodology

The original authors only briefly describe their second study, on pages 521-521 of their paper.



Participants were presented with pairs of summaries, and asked to choose the better summary of the two in terms of (definitions cited from the original paper):

- **Grammaticality:** *is the summary written in well- formed English?*
- **Coherence:** *is the summary well structured and well organized and does it have a natural ordering of the facts?*
- **Conciseness:** *does the summary avoid unnecessary repetition including whole sentences, facts or phrases?*

The original authors note that they “divided the five competing systems into ten pairs of summaries and elicited ratings for 40 summaries (20 per dataset). Each summary pair was rated by 3 raters. This resulted in 40 summaries × 10 system pairs × 3 evaluation criteria × 3 raters, for a total of 3,600 tasks.” These tasks were uploaded to the Mechanical Turk platform to be rated by crowd-workers. The authors reported 206 workers contributing to this process, with different workers providing different amounts of ratings.

We found an error in the original data, meaning that we cannot use half of the data, so we will use 20 summaries × 10 system pairs × 3 evaluation criteria × 3 raters, for a total of 1,800 tasks. Presumably this will require about 100 workers to provide ratings, but since workers can provide any amount of ratings, this number may differ.

We will analyse the data using the same methods as the original authors.

#### 1.4 Scientific and societal relevance

This project enables us to understand better to what extent existing human evaluation studies can be reproduced. This particular study provides one of the data points in the project. Besides that, it helps us as a research team better understand the difficulties of reproducing existing research (which helps us in the reporting of future studies), and it tells us more about the results from the study that we are reproducing.

The societal relevance is less direct, but this kind of fundamental research is needed to make our reporting more reliable, and thus more useful for people building on our work.



### 1.5 How will data be collected?

Please note that data minimization is the norm: do not collect more data than is necessary for answering the research question(s).

- ☐ Structured individual interviews
- ☐ Semi-structured individual interviews
- ☐ Structured group interviews
- ☐ Semi-structured group interviews
- ☐ Observations
- ☐ Literature study
- ☐ Survey(s)
- ☐ Lab experiment(s)
- ☐ Experiment(s) in real life (interventions)
- ☐ Physiological data, eye tracking
- ☐ Secondary analyses on existing datasets
- ☒ Other:

Online crowdsourcing experiment, similar to a survey.

### 1.6 Which tools will be used for data collection and/or data analyzing?

#### ☒ Online

- ☒ Mturk
- ☐ Qualtrics
- ☐ Survey Monkey
- ☐ Other, please specify:  
[Click here to enter text.](#)

- ☐ Paper-and-pencil
- ☐ Recordings: phone, photo, audio or video
- ☐ Via a mobile application or wearables, please specify:  
[Click here to enter text.](#)

- ☐ Face-to-face
- ☐ Within an online group
- ☐ Within a real life group
- ☐ Other:  
[Click here to enter text.](#)



## Part B: Ethics

---

### 2. PARTICIPANTS

#### 2.1 Please check the box that indicates the relevant study population:

- ☐ Students
- ☒ General population
- ☒ Specific groups: crowd-workers from English speaking countries, including US, UK, Canada, NZ, Aus and Ireland.

#### 2.2 Age category of the participants:

- ☐ Younger than 12 years of age
- ☐ Older than 11 and younger than 16 years of age
- ☐ Older than 15 and younger than 18 years of age
- ☒ 18 years and older

#### 2.3a Method of recruitment or selection of participants:

Participants will be recruited by posting our tasks on the MTurk website. We will provide a fair compensation based on the local minimum wage, through a formula provided by the University of Aberdeen.

#### 2.3b State the organization where the recruitment of participants will take place:

- ☐ Tilburg University
- ☒ Other, namely: Mechanical Turk through the account of the University of Aberdeen (due to financial reasons).
- ☐ Not applicable because: [Click here to enter text.](#)

#### 2.4 Reward for participation (multiple answers are possible):

- ☐ None
- ☐ Reimbursement of (travel) expenses
- ☐ Course credit
- ☒ Financial reward, i.e. At or above local minimum wage. See above (2.3a) €/hours
- ☐ Other, namely: [Click here to enter text.](#)

#### 2.5 Describe in detail the expected burden and/or potential negative consequences for the participants with respect to time, mental, and/or physical burden. Pay special attention to any kind of 'deception'.

No deception takes place. The task might be a bit boring/tedious to participants, but otherwise without any negative consequences. It's just a task where participants rate the quality of summaries of sports events.



**2.6 Describe measures that have been taken to protect the participant (e.g. additional insurance, debriefing, etc.):**

Information letter and informed consent are sufficient to inform the participant about the nature of the task, but since there are no risks, there is no need to further protect the participants.

☐ Not applicable, because: [Click here to enter text.](#)

**2.7 Are participants subjected to procedures or experiment-related manipulations or tasks? Indicate which ones, how, and with what purpose.**

Yes, as in the original study: participants are presented with two summaries, provided by different systems, to determine which system produces the best summary.

**3. ADDITIONAL INFORMATION**

Please use this space to ask any additional questions you might have for the REDC and/or to add information that is relevant to your project, but that is not asked in this form.

This project is similar to earlier applications by Anouck Braggaar and Chris van der Lee, with only minor differences. See REDC 2019.40abc for our previous replication study.

**4. CHECKLIST INFORMATION TO PARTICIPANT**

*Please check each applicable box to confirm that the information letter contains the required elements*

- ☒ Title (Title of the study, if necessary simplified, abbreviated or translated)
- ☒ Introduction

**4.1 What does the study entail?**

- ☒ Purpose
- ☒ Background
- ☒ Nature
- ☒ Duration

**4.2 What does participating in the study entail?**

- ☒ Procedures
- ☒ Expected duration
- ☒ Disadvantages/consequences/risks
- ☒ Possible advantage for the participant

**4.3 Information about the participation**

- ☒ Voluntariness of participation.
- ☒ Right to decline to participate and withdraw from the research once participation has begun, without any negative consequences, and without providing any explanation.
- ☐ Right, in principle, to request access to and rectification, erasure, restriction of or object to the processing of the personal data. For more information: [www.tilburguniversity.edu/privacy](http://www.tilburguniversity.edu/privacy)

**Commented [EM1]:** No personal data will be collected.



- ☒ Confidentiality protection and limitations
- ☐ Applicable insurance guarantees (only if there is additional insurance to the standard insurance)
- ☒ Period of time to which the consent applies (normally the length of the study)
- ☒ Re-use of specified data in the current, future or other research, where applicable
- ☒ Deliberation time (if applicable)
- ☒ How the data will be processed
- ☒ Period of time that data will be [stored and encrypted](#)
- ☒ Incentives for participation (traveling expense, pp hours)
- ☒ Approval by Research Ethics and Data Management Committee (REDC)
- ☒ Request for participation
- ☒ The following text should be included:

Voor eventuele opmerkingen of klachten over dit onderzoek kunt u ook contact opnemen met de "Research Ethics and Data Management Committee" van Tilburg School of Humanities and Digital Sciences via [tshd.redc@tilburguniversity.edu](mailto:tshd.redc@tilburguniversity.edu)

If you have any remarks or complaints regarding this research, you may also contact the "Research Ethics and Data Management Committee" of Tilburg School of Humanities and Digital Sciences via [tshd.redc@tilburguniversity.edu](mailto:tshd.redc@tilburguniversity.edu)
- ☒ Closing / whom to contact in case of question or additional information (name and telephone number/ email address researchers)
- ☒ Appendices: Informed Consent

## 5. CHECKLIST INFORMED CONSENT

*Please note that the consent form should be written in a language that is understandable to participants and that it should not be reduced to a list of statements or a series of check boxes. Its purpose is to inform prospective participants of the purpose, procedures, and risks and benefits of participating in the study, so that they can make an informed decision about whether they want to participate, free of explicit or perceived coercion.*

- *In case of minors younger than 12 years of age informed consent is obtained from the parent(s) or legal representative(s). It is default to also ask the child where possible.*
- *In case of minors older than 11 and younger than 16 years of age informed consent is obtained from both the minor and the parent(s) or legal representative(s).*
- *From 16 years of age, consent is only obtained from the participant. For some types of research it may nevertheless be default to inform the parents or legal representatives.*
- *In case of a mentally incompetent participant, informed consent is obtained from the legal representative(s). It is default to also ask the participant where possible.*

*Please check each applicable box to confirm that the informed consent contains the required elements*

### 5.1 Mentally competent participants, including minors between the age of 12 and 16 year

- ☒ Title (Title of the study, if necessary simplified, abbreviated or translated)
- ☒ Confirmation that the information is read
- ☒ Confirmation that there was room for questions
- ☒ Reminder on voluntariness of participation. Right to decline to participate and withdraw from the research once participation has begun, without any negative consequences, and without providing any explanation



- ☒ Permission processing of anonymous/coded data as mentioned in the information letter
- ☒ If applicable: explicit permission to make photos, audio/video recordings
- ☒ Permission for storing the research data for a period of at least ten years
- ☒ Permission participation in the study
- ☐ Date, name, signature participant
- ☐ Date, name, signature researcher
- ☐ Give the participant a copy of the signed informed consent form

Commented [EM2]: Not applicable; online form.

#### 5.2 Addition/correction for minors

- ☐ Date of birth participant
- ☐ Date, name, signature (if possible both) parents/guardians

#### 5.3 Addition/correction for mentally incompetent adults

- ☐ Date, name, signature legal representative, relation to participant

#### 6. Documents that should be added to the application

The following documents should be provided (if applicable):

- ☐ Advertisement
- ☒ Participant information letter (precedes participation)
- ☒ Informed consent form
- ☐ Written debriefing
- ☒ All surveys/questionnaires that will be used
- ☒ Description of the stimulus materials
- ☐ Written consent of organization(s) (except Tilburg University) to recruit participants

## Part C: Data Management – Data storage

The questions in this section form, together with some questions in section A, the Data Management Plan (DMP) for your research. Please, fill in this part completely regardless of the type of study you are performing. The DMP is





obligatory for every research project. For more information check the [Tips for writing a Data Management Plan](#) or contact the [Research Data Office](#).

## 7.1 Data storage

- Storage locations are the digital locations where you store your data: allowed are the TiU network drives (M, O, P) DataverseNL, SharePoint, Surfdrive, Research Drive (if using other locations, please explain). You can check [here](#) which location is suitable for your type of data. Please note that sensitive data need to be protected (encryption, locked cabinet, etc.).
- Access to raw data. The raw database is the first database obtained digitally by the staff member. Ideally this should be first digital file created by anyone, but in cases where datasets/secondary datasets from third parties or files e.g. downloaded from data repositories or other types of publicly available databases are used this file may also mean these. Self-collected data means data such as entered questionnaires or data collected by means of online surveys, computers or measuring tools. Please indicate who has access to the raw data and in what time periods. Please mention both the names and the role of the persons who has access, for example: [name], principal investigator, [name], head of department, [name], research assistant, etc.
- Access to processed data, "research ready" data which has been fully calibrated, combined and cleaned/annotated. Please indicate who has access to the processed data and in what time periods. Please mention both the names and the role of the persons who has access, for example: [name], principal investigator, [name], head of department, [name], research assistant, etc.
- Storage/archiving period: the period of time the data package is securely stored and saved in a (digital) repository.
- Data collection period: the period of time in which all data is gathered to perform the research.
- Data analysis period: the period of time in which all data is analyzed to perform the research.
- Data archiving: the period of time after the study has finished and data is stored in a secure (digital) facility.

	Data collection period	Data analysis period	Data archiving
Storage location	Mechanical Turk	SurfDrive, work laptop	GitHub (Code+data)
Access to raw data	PI (Emiel van Miltenburg) Account provider (Craig Thomson, project coordinator at University of Aberdeen)	All local collaborators. (Except potentially identifying information such as IP address is removed)	Everyone (public data) (Except potentially identifying information such as IP address is removed)
Access to processed data	PI (Emiel van Miltenburg and all local collaborators)	All local collaborators	Everyone (public data)
Storage/archiving period (years)			<u>Indefinitely</u>

**7.2 Where will data be preserved long-term (for example a data repository or archive)?** Tip: [Tilburg University Dataverse](#) is a data repository managed by LIS Research Data Office and is available for all TiU researchers.



- ☒ Dataverse
- ☐ O-drive
- ☒ Other: GitHub

**7.3 Which criteria will you use to decide which data has to be archived for preservation and long-term access? Which (part of the) data has to be destroyed to ensure privacy protection? For guidance see part 5 of the [Tips for writing a data management plan](#).**

Everything is fully anonymous, so nothing needs to be destroyed. All files are relevant, particularly since this is a reproduction study.

#### 7.4 Meta data

Metadata are data about your data and are used to describe your data set/package. They can be descriptive (common fields such as title, author, abstract, keywords which help users to discover online sources through searching and browsing), administrative (preservation, rights management, and technical metadata about formats), or structural (how different components of a set of associated data relate to one another, such as a schema describing relations between tables in a database). A metadata standard is a structured way of describing data. Repositories often use an existing standard. For TiU Dataverse this is the DDI (Data Documentation Initiative) standard.

Please indicate what will be included in your metadata, how it will be documented, and if you use a metadata standard and which one (for guidance see part 7 of the [Tips for writing a data management plan](#)).

We will write an extensive README document explaining the files in the project folder, and the software requirements to run our code. Each file will have a descriptive name. The GitHub repository will be linked from the research paper. Since the data is not very complex (it is an Excel file with human ratings), this is sufficient for any potential reuse.

#### **In case of non-digital data storage of the main data of a study**

**7.5 What kind of non-digital data will need to be stored during the study (paper surveys, transcripts, informed consent forms, photocopies of original documents, ...)? Do these data need special protection to secure privacy and confidentiality? Where will these data be stored? Who will have access to the data?**

N/A

**7.6 What kind of data will need to be stored over the required term of 10 years, for long-term preservation of data? For example: processed, anonymized datasets which do not have to be deleted.**

N/A

## Part C: Data Management - Data sharing

---

**7.7a) Will (part of) the data be made available for reuse after completing the project according to the FAIR [Findable, Accessible, Interoperable, Reusable] Principles?**

For guidance see part 6 of the [Tips for writing a data management plan](#). If so, please describe in a concrete manner when and how the data will be made available (see part 11 of the [Tips for writing a data management plan](#)).

Yes

**b) Are there possible restrictions to data sharing or embargo reasons? Are there any conditions for the re-use of the data? If so, which conditions and how will the data be shared? What license will be applied to your research data?**

For guidance see <http://www.dcc.ac.uk/resources/how-guides/license-research-data> Tip: data repositories often provide licenses to choose from.

No restrictions. Code and data is provided under MIT/CC-BY licenses.

**c) In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).**

N/A

**7.8 How will ownership of the data and intellectual property rights to the data be managed?**

Explain who will be the owner of the data, meaning who will have the rights to control access.

We own the data, but make everything freely available for both commercial and non-commercial use.



## Part D: GDPR / Data Processing Register

---

By filling out this part of the form you are complying with the GDPR which requires personal data to be included in the data processing register of Tilburg University. This includes a pre-DPIA (Data Protection Impact Assessment), which makes it visible whether there are certain risks and whether you are obliged to conduct a DPIA.

### 8.1 Which personal data are to be collected and processed?

- ☐ No personal data will be processed. This is the end of the questionnaire, you can skip question 8.2 until 8.9.
- ☒ Yes, namely (multiple answers possible):

#### General

- ☐ Contact data (for example: name, email address, phone)
- ☐ Gender
- ☐ Age, birthdate
- ☐ Nationality, birthplace, birth country
- ☐ Student number/employee number
- ☐ Experiences (work, education)
- ☐ Finances
- ☐ Visual materials (pictures, video)
- ☒ IP-address

#### Special data

- ☐ Racial or ethnic origin
- ☐ Religious or philosophical beliefs
- ☐ Political opinions
- ☐ Health data
- ☐ Sex life or sexual orientation
- ☐ Trade union membership
- ☐ Genetic data
- ☐ Medical data
- ☐ Biometric data
- ☐ Criminal records

#### Sensitive data

- ☐ Copy identification card
- ☐ R&D meeting
- ☐ Study results

#### Other

- ☐ Namely, [Click here to enter text.](#)



**8.2 Will data be anonymized or pseudonymized after collection and if so, who will have access to the identifying file? Please explain. Are additional measures necessary in case of audio/video recordings or other kind of recordings? Please explain.**

Yes. IP addresses will be removed or pseudonymized. If Amazon still provides worker IDs, then these will be pseudonymized as well.

**8.3 Are there any external parties (processors) involved in this study regarding data collection, data storage, archiving and/or other data-related activities? If so, please describe and name them here and state the website(s) of the processor(s) and / or external controller(s).**

A processor is a person or organization to whom or which the responsible party has outsourced the processing of personal data, such as a cloud service. The external party should comply with the GDPR. Some services, such as Surfdrive, comply with these regulations. If an external party is not known to be GDPR compliant, the applicant should ensure that there is a contract to ensure appropriate processing by the external party. This party should take appropriate technical and organizational measures to protect personal data against loss or any form of unlawful processing (e.g. unnecessary collection of data or further processing).

The model processor agreement and procedure is available via [intranet](#) or can be requested from one of the Data Representatives.

**Data collection**

☐ Not applicable

☒ Yes: Mechanical Turk.

**Data storage**

☒ Not applicable

☐ Yes: [Click here to enter text.](#)

**Data archiving**

☒ Not applicable

☐ Yes: [Click here to enter text.](#)

**Other data-related activities (e.g. analyses)**

☒ Not applicable

☐ Yes: [Click here to enter text.](#)

**Only applicable if there is an external party that requires a processor agreement**

**8.4 Have you agreed upon and centrally archived a processor agreement with the above-mentioned processors? Please specify.**

[Click here to enter text.](#)

**Commented [EM3]:** As far as we know it is compliant. See: [https://www.amazon.com/gp/help/customer/display.html/ref=footer\\_privacy?ie=UTF8&nodeId=468496#GUID-8966E75F-9B92-4A2B-BFD5-967D57513A40\\_SECTION\\_A110DAC3F6BC4D5D9DDD59797104B1E5](https://www.amazon.com/gp/help/customer/display.html/ref=footer_privacy?ie=UTF8&nodeId=468496#GUID-8966E75F-9B92-4A2B-BFD5-967D57513A40_SECTION_A110DAC3F6BC4D5D9DDD59797104B1E5)

**Commented [EM4]:** At this point the data is anonymous, so no PII is present.



#### 8.5 What is the legal base for which the processing activity takes place?

Personal data cannot be processed according to the GDPR unless there is a legal base for processing. This question refers to the legal basis for processing personal data. There are several possibilities, three of which are particularly relevant for research at Tilburg University. Indicate the option applicable to your project:

- ☒ (1) Consent (participants sign a consent form to process their personal data);
- ☒ (2) Legitimate interest as scientific researcher ("gerechtvaardigd belang");
- ☒ (3) Permission (when an external party provides the applicant with personal data and the external party has obtained consent to use these data)

#### 8.6 Does the applicant receive personal data from or provide personal data to a third party and which of the organizations determines the purpose and means of the processing?

This happens when an applicant receives data from an external party and the external party determines what will happen with the data (e.g., the applicant receives data on which specific analyses have to be conducted). Then it is mandatory to make specific agreements regarding the delineation of the processing. The model processor agreement and procedure is available via [intranet](#) or can be requested from one of the Data Representatives

☒ No

☐ Yes, data will be sent to:

- ☐ The project group, including [Click here to enter text.](#)
- ☐ Co-researcher from other universities of institutions. Please state their names, contact details and countries: [Click here to enter text.](#)
- ☐ Other persons responsible for processing the data. Please state their names, contact details and countries: [Click here to enter text.](#)

☐ Yes, data access will be provided to:

- ☐ The project group, including [Click here to enter text.](#)
- ☐ Co-researcher from other universities of institutions. Please state their names, contact details and countries: [Click here to enter text.](#)
- ☐ Other persons responsible for processing the data. Please state their names, contact details and countries: [Click here to enter text.](#)

#### 8.7 If applicable, to which third parties (controllers and processors) are the data provided by default? What is the purpose and the basis of this provision?

Examples are tax authorities, pension funds, health insurers etc. Third parties with an independent processing responsibility are always external and determine their own purpose and resources for the processing. If the data is provided to another controller, then an agreement should be concluded about privacy and security guarantees. This



can be done in the agreement that already exists with that other party or in a data exchange agreement for the study for which this clearance is asked.

N/A

**8.8 Is a Data Protection Impact Assessment (DPIA) needed? Tick all categories<sup>1</sup> that apply to your research. If a DPIA is required, the Data Representative will contact you to schedule this.<sup>2</sup>**

A DPIA is an estimate of the impact of data processing on the data protection of the persons concerned. Such an assessment is required if the intended processing of personal data poses a high privacy risk to the persons concerned, for example if the applicant intends to collect a huge data set or an extremely sensitive data set. Such processing warrants a separate analysis of the risks of the project. Based on this estimate, recommendations can be made to minimize this impact as much as possible or even eliminate it completely.

☐ **Assessing people on the basis of personal characteristics:** this includes profiling and predicting, particularly on the basis of characteristics such as a person's professional performance, economic situation, health, personal preferences or interests, reliability or behavior, location, or movements. Examples include a bank that determines the creditworthiness of customers (credit scoring), a company that provides DNA tests to consumers to test health risks, and a company that follows visitors to its website and uses this to create profiles of these people.<sup>3</sup>

☐ **Automated decisions:** this concerns making decisions with technological means and without human intervention. In order to fall within this category, the decisions should have legal effects or comparable significant effects on the person concerned. Such data processing may, for example, lead to exclusion or discrimination. Data processing with little or no impact on individuals is not covered by this criterion.

☐ **Systematic and large-scale monitoring:** this concerns the monitoring of publicly accessible spaces, for example with camera surveillance, but also systematically following data subjects online. Personal data can be collected without those involved knowing who is collecting their data and what happens to it. Additionally, it may be impossible for people to withdraw from this data processing in public places.

☐ **Sensitive data:** this concerns special categories of personal data, as laid down in Article 9 of the GDPR: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. Moreover, this category also includes

<sup>1</sup> The European Working Party 29 (WP29) has indicated nine criteria for which, if there are at least two applicable, a DPIA must be carried out. These criteria are endorsed by the European Data Protection Board and the Dutch Data Protection Agency ("Autoriteit Persoonsgegevens"). Tilburg University has added the criterion about data transfer across borders of the EU due to special regulations that apply in that situation. The criteria are written mostly for large corporations processing personal data and do not take the specifics of scientific research into account. An additional explanation by Tilburg University is given for some of the criteria for scientific research.

<sup>2</sup> Please note that as a rule of thumb a DPIA is required when two or more categories apply. However, it is possible that a DPIA is required if one or even none of the criteria are applicable.

<sup>3</sup> At Tilburg University research is conducted in which characteristics of individuals might be used to segment individuals into different groups to explain for example their behavior. This can be seen as profiling. If the processing is done for research purposes and it does not affect the individuals directly, this does not present a high level of privacy risk, which is why in those cases, the criterion **will not** be applicable. However, if the research is conducted as for example contract research and the results directly affect individuals, this criterion **will be** applicable.



data that are generally regarded as privacy sensitive, such as data about electronic communication, location data and financial data.

☐ **Large-scale data processing:** there is no specific definition of this category, but the following criteria should be used to determine whether this is applicable: a. very large datasets concerning many thousands or millions of people; b. the volume of data and/or the range of different data items being processed; c. the duration of the data processing activity; d. the geographical extent of the processing activity.<sup>4</sup>

☐ **Combining databases:** Datasets that have been matched or combined, for example originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would make it possible to deduce the personal identities of subjects.

☐ **Data concerning vulnerable data subjects:** this applies when there is an unequal balance of power between the data subject and the data controller for example minors, mentally ill persons, asylum seekers, or the elderly, patients, etc.

☐ **Use of new technologies:** e.g., combining use of fingerprint and face recognition for improved physical access control, etc. The reason is that this use may involve new ways of collecting and using data, with potentially high privacy risks. The personal and social consequences of using a new technology may even be unknown, a DPIA then helps to understand and remedy the risks.

☐ **Data transfer across borders outside the European Union,** taking into consideration, the potential risks of data transfers to such countries.

☐ **Blocking of a right, service, or contract:** this concerns data processing that result in data subjects not being able to exercise a right, use a service, or enter into a contract.<sup>5</sup>

<sup>4</sup> The Dutch Data Protection Agency has determined that for the health care industry 10,000+ individuals make up a large dataset. For other industries, no number has been provided. However, depending on the type of data and the number of data points per individual, a smaller number of individuals might make up a large dataset because the processing of the data will likely present a **high level of privacy risk**. When in doubt, check with your data representative.

<sup>5</sup> This criterion is applicable for example when covert research is conducted since data subjects are not aware they are being part of scientific research and can therefore not claim their right to information, to object to the processing of their information, etc.