

Begeleidende ethische vragen

In dit document vind je (voorbeelden van) begeleidende ethische vragen in de context van werken met data. Het doel van dit document is bewustzijn creëren over de ethiek van werken met data.

Dit document heeft de volgende uitgangspunten:

- De vragen zijn ingedeeld per fase van een project. Omdat projecten verschillende vormen aannemen, bepaal je zelf welke vragen van toepassing zijn op jouw project.
- De lijst is niet volledig en zal regelmatig worden bijgewerkt.
- Het doel van de lijst is niet om vragen te beantwoorden, maar discussie te stimuleren.
- De vervolgvraag op elke vraag is: “en is dit ethisch verantwoord?”

1 Conceptualisation: een concept of idee van iets maken

- Wat is het doel van een project?
- Wat zijn de positieve en negatieve effecten op stakeholders?
- Welke effecten zijn belangrijker dan anderen en waarom?

Om bovenstaande vragen te beantwoorden, moeten we eerst weten:

- Wie zijn de stakeholders? Maak onderscheid tussen interne stakeholders (marketeers, analisten, management) en externe stakeholders (klanten, indirecte eindgebruikers).
- Wat zijn potentiële voordelen van een project voor <stakeholders>? Welke ethische waarden worden aangehaald door het project (gemak, veiligheid, relevantie, efficiëntie, betrouwbaarheid)?
- Wat zijn potentiële nadelen van een project voor <stakeholders>? Wat zijn ethische waarden die negatief beïnvloed worden door het project (veiligheid, gelijkheid, autonomie, privacy)?

De waarden die hierboven genoemd worden, vormen de basis van alle vervolgstappen in het project. Bijvoorbeeld: als (sample) data een bias bevat, wat zijn dan de effecten op de waarden van verschillende stakeholders? Voorbeelden van zulke effecten zijn:

- Als het profiel van een klant gebaseerd is op biased historische data, kan het zijn dat een klant verdacht wordt van fraude (zoals gebeurde bij de Belastingdienst).
- Als een klant verkeerd geprofileerd wordt, kan zij/hij een verkeerde aanbieding ontvangen en zich hierdoor gediscrimineerd voelen (in het geval van e-commerce organisaties).
- Als de data van een klant (met initiële toestemming) gebruikt wordt voor andere doeleinden dan in eerste instantie besproken, wordt de privacy van de klant geschonden en staat de reputatie van een organisatie op het spel.

2 Operationalisation: de meetmethode van een concept of idee definiëren

- Waar is de selectie van een sample op gebaseerd/wat is de sampling strategie?
- Is een geselecteerd sample voldoende om de vraag/vragen te beantwoorden?
- Heeft de sampling strategie biases?
- Zijn er beperkingen van de gebruikte meetmethode en -instrumenten (reliability)?
- Kun je meten wat je moet/wilt meten (validity)?

3 Data verzamelen

Voor primaire datacollectie:

- Gebruiken we persoonlijke data alleen voor het doel en gedurende de periode die we met data subjects hebben gecommuniceerd? (=informed consent, verplicht onder AVG)
- Is datacollectie beperkt tot alleen de (persoonlijke) data die nodig is voor het beantwoorden van een vraag of voor het doel van een organisatie (gerelateerd aan de klantreis)?
- Zijn er biases te vinden in de verzamelde data?
- Is er een retentieperiode gedocumenteerd voor de data?

4 Data analyseren

- Zijn er biases in de analysemethodes terug te vinden? Zo ja: is dit een probleem?
- Zijn er alternatieve analysemethodes overwogen? Zo ja: waarom zijn deze niet gebruikt?
- Hoe wordt er met outliers omgegaan (zijn outliers errors of zijn ze ongewenst)?
- Worden we aangemoedigd om een onafhankelijke analyse te doen? Of staan we onder druk om bepaalde (bijvoorbeeld positieve) resultaten te rapporteren?

5 Data modelleren

- Bevatten de modellen (significante) biases? Zo ja: is dit een probleem?
- Bevatten de resultaten (significante) biases? Zo ja: is dit een probleem?
- Zijn de resultaten van een project bruikbaar voor de eindgebruiker?
- Hebben de resultaten significante sociale consequenties? (Bijvoorbeeld wanneer processen geautomatiseerd worden, wat zijn dan de effecten op de organisatie en eindgebruikers?)
- Hoe wordt er met outliers omgegaan?
- Is er iemand aansprakelijk en bevoegd om vragen en klachten van eindgebruikers of collega's te beantwoorden?

6 Data visualiseren

- Zijn biases in het sample/datacollectie/analyse/modelleren of andere biases voldoende gevisualiseerd?
- Zijn de visualisaties een eerlijke representatie van de onderzoeksresultaten?
- Zijn de visualisaties begrijpelijk voor de doelgroep?
- Zijn er geen 'verstopte' resultaten die het publiek niet zal begrijpen?
- Worden verschillende helder genoeg gevisualiseerd (bijvoorbeeld in kleurcodes)?
- Wordt de N gerapporteerd bij de visualisaties?
- Zijn sub-samples gerapporteerd bij de visualisaties (bijvoorbeeld Nederlandse vrouwen)?

- Zijn databronnen gerapporteerd bij de visualisaties?

7 Data rapporteren

- Zijn biases in het sample/datacollectie/analyse/modelleren of andere biases voldoende besproken?
- Is er voldoende aandacht besteed aan het uitleggen van de beperkingen van een project aan verschillende stakeholders?
- Zijn de nodige stappen genomen om misinterpretatie van een project en de resultaten te voorkomen?
- Gebruikt een rapport de juiste schalen?
- Zijn de resultaten juist geïnterpreteerd binnen de context van een project?
- Zijn vergelijkingen gemaakt met een juiste nulmeting?
- Is de N gerapporteerd?
- Concludeert het rapport terecht causaliteit of correlatie?

8 Data beheren

Voor alle data types (persoonlijk en niet-persoonlijk):

Organisatie

- Handelen stakeholders naar ethische overwegingen?
- Staat de structuur van een organisatie ethisch gedrag toe? Wordt het gestimuleerd?
- Zijn er tegenstrijdige belangen binnen de organisatie die een effect hebben op databeheer?
- Hebben stakeholders hetzelfde begrip van alle datapunten? (Verschillende definities kan leiden tot misinterpretatie en negatieve effecten op de waarden van stakeholders)
- Zijn KPI's duidelijk?

Toegang en opslag

- Hebben alleen degenen die direct met data moeten werken toegang tot de data? Anderen hebben waarschijnlijk voldoende aan inzichten uit de data.
- Hebben alleen degenen die resultaten juist kunnen interpreteren toegang tot de data?
- Kan niet-persoonlijke data worden herleid naar individuen (ook door slimme mensen die behoorlijk hun best doen)? Zo ja: zijn de nodige stappen ondernomen om risico's voor data subjects te minimaliseren?
- Is datacollectie goed gedocumenteerd en bijgewerkt?
- Is er een opslagbeleid en is dit goed gedocumenteerd?
- Als data op servers van derden wordt opgeslagen: zijn de compliance policies van deze partijen bekend?
- Als data op servers van derden wordt opgeslagen: hebben deze partijen toegang tot de data?
- Kan de retentieperiode van data verkort worden naar minder dan de maximaal toegestane 12 maanden (volgens de AVG)?
- Is er een proces voor data verwijdering (inclusief welk type data wordt verwijderd, hoe het wordt verwijderd, wanneer en door wie)?

Specifiek voor persoonsgegevens

- Is persoonlijke data duidelijk gelabeld?
- Is er een proces om data te verwijderen wanneer eindgebruikers hierom vragen? (= recht op vergetelheid, AVG)
- Is er een proces om eindgebruikers toegang te geven tot hun eigen data? Of om data aan te passen?
- Zijn bronnen of systemen die persoonlijke data bevatten alleen aan elkaar gekoppeld wanneer dit is toegestaan volgens de AVG?