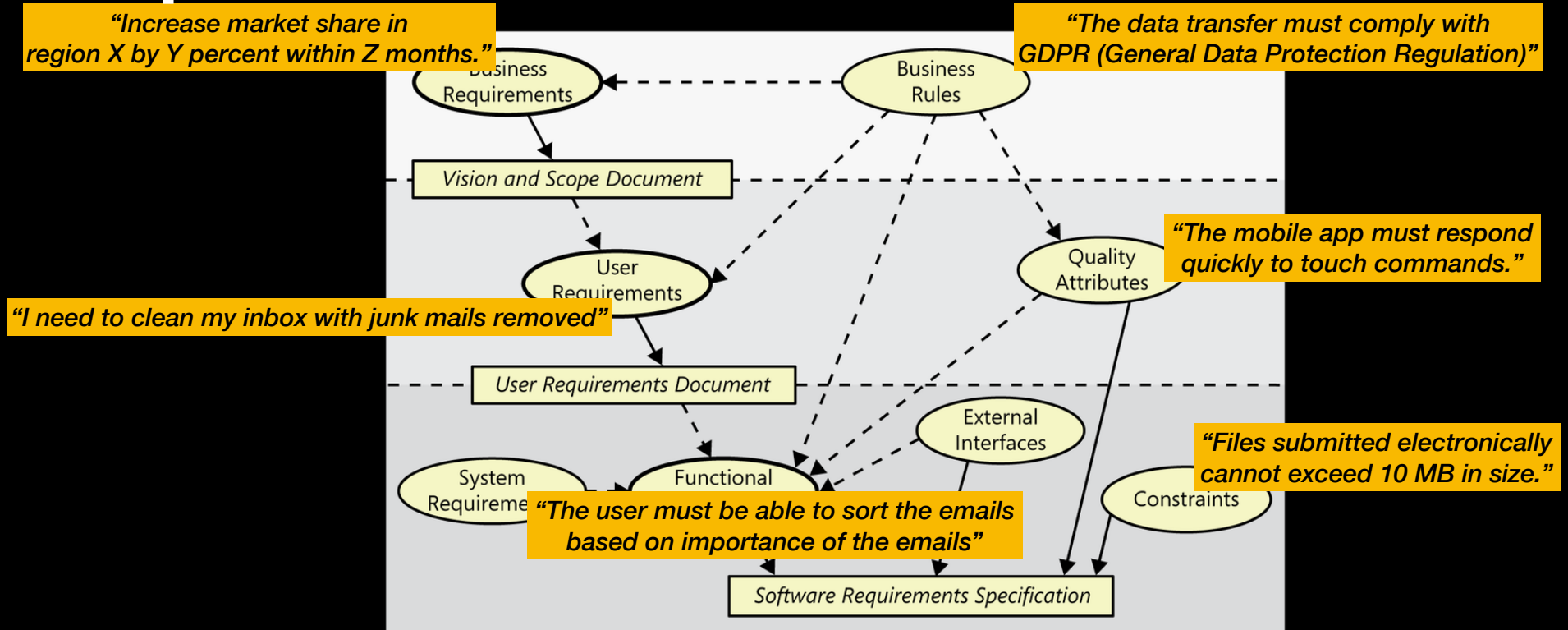# Requirement & Intelligent Systems

Jin Guo   Oct 8th, 2020

Requirement Definition

- Requirements are a specification of what should be implemented.

- They are descriptions of how the system should behave,

- or of a system property or attribute.

- They may be a constraint on the development process of the system.

Sommerville, Ian, and Pete Sawyer. 1997. *Requirements Engineering: A Good Practice Guide.* Chichester, England: John Wiley & Sons Ltd.

Requirement Definition

"Increase market share in region X by Y percent within Z months."

"The data transfer must comply with GDPR (General Data Protection Regulation)"

"The mobile app must respond quickly to touch commands."

"I need to clean my inbox with junk mails removed"

"Files submitted electronically cannot exceed 10 MB in size."

"The user must be able to sort the emails based on importance of the emails"

Diagram nodes: Business Requirements, Business Rules, Vision and Scope Document, User Requirements, Quality Attributes, User Requirements Document, External Interfaces, System Requirements, Functional, Constraints, Software Requirements Specification

Software Requirements by Karl Eugene Wiegers, Joy Beatty, Microsoft Press, 2013

Example Requirement Specification
"Global Personal Marketplace SRS"

- How are the Functional Requirements (FRs) organized in this example?

- What kind of information is included in the FRs?

Data Requirements

"An order consists of the customer's identity, shipping information, and one or more products, each of which includes the product number, number of units, unit price, and total price."
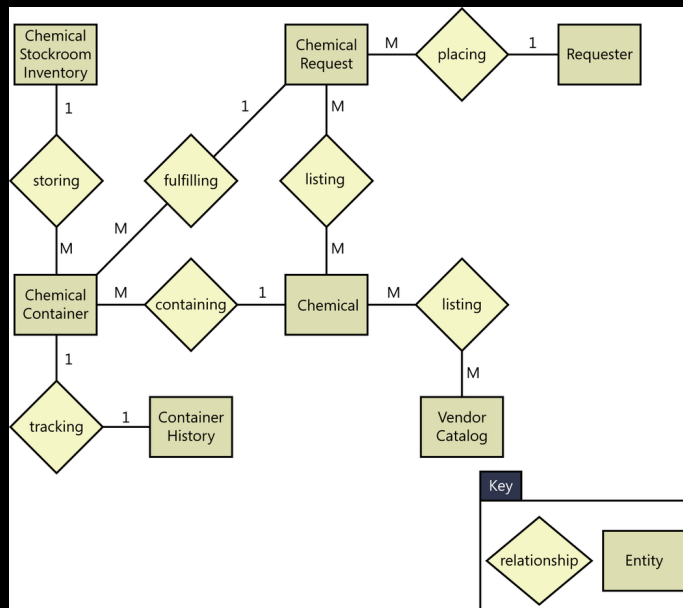
Data Element	Description	Composition or Data Type	Length	Value
Chemical Request	*request for a new chemical from either the Chemical Stockroom or a vendor*	Request ID + Requester + Request Date + Charge Number + 1:10{Requested Chemical}		
Delivery Location	*the place to which requested chemicals are to be delivered*	Building + Lab Number + Lab Partition		
Number of Containers	*number of containers of a given chemical and size being requested*	Positive integer	3	
Quantity	*amount of chemical in the requested container*	numeric	6	
Quantity Units	*units associated with the quantity of chemical requested*	alphabetic characters	10	grams, kilograms, milligrams, each
Request ID	*unique identifier for a request*	integer	8	system-generated sequential integer, beginning with 1
…	…	…	…	…

Software Requirements by Karl Eugene Wiegers, Joy Beatty, Microsoft Press, 2013

User stories

As a <type of user>, I want <some goal> so that <some reason>.

Use cases

• Update Customer Profile

• Search for an Item

• Buy an Item

• Track a Shipped Package

• Cancel an Unshipped Order

As a customer, I want to update my customer profile so that future purchases are billed to a new credit card number.

Feature

- A feature consists of one or more logically related system capabilities that provide value to a user and are described by a set of functional requirements.
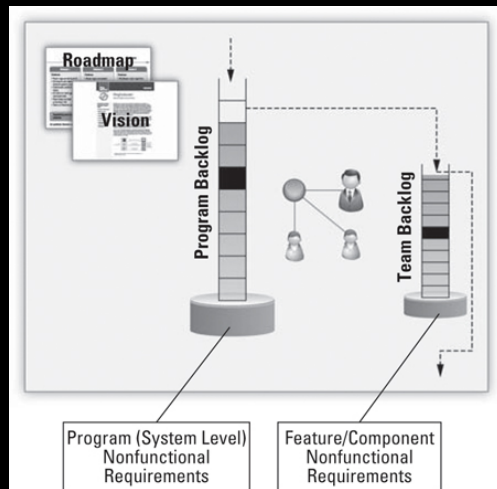
Agile Software Requirements Lean Requirements Practices for Teams, Programs, and the Enterprise, by Dean Leffingwell

Elicitation

Identifying the product's expected user classes and other stakeholders. Understanding user tasks and goals and the business objectives with which those tasks align.

Validation

Reviewing the documented requirements to correct any problems before the development group accepts them. Developing acceptance tests and criteria to confirm that a product based on the requirements would meet customer needs and achieve the business objectives.

Analysis

Decomposing high-level requirements into an appropriate level of detail; Deriving functional requirements from other requirements information; Understanding the relative importance of quality attributes; Allocating requirements to software components defined in the system architecture Negotiating implementation priorities

Specification

Translating the collected user needs into written requirements and diagrams suitable for comprehension, review, and use by their intended audiences.

Challenges
Insufficient User Evolvement

- Users vs Stakeholders

    - Examples? What are the differences?

- Classifying users

    - Access privilege or security level

    - The tasks they perform

    - The features they user

    - Domain expertise

    - ...

Are different classes of users equally important for the system you are building?

?

Classifying users

- Favoured user classes

  Receive preferential treatment when resolving conflicts between requirements
  from different user classes or making priority decisions

- Disfavoured user classes

  Might deliberately make it hard for the disfavoured users to do things they aren't supposed to do

- Indirect user classes

  Access its data or services through other applications or through reports

User classes need not be human beings.

Activity

- Who are the user classes for the Github pull request (PR) management system?

  - Example PR1, PR2

Indirect Stakeholders

"Janice is in her office, writing a report. She's trying to conceptualize the report's higher-level structure, but her ideas won't quite take form. Then she looks up from her desk and rests her eyes on the fountain and plaza area outside her building. She notices the water bursting upward, and that a small group of people are gathering by the water's edge. She rests her eyes on the surrounding pool of calm water. Her eyes then lift toward the clouds and the streaking sunshine. Twenty seconds later she returns to her writing task at hand, slightly refreshed, and with an idea taking shape."

Figure 2: Plasma Display Technology Studies

Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value sensitive design and information systems." *The handbook of information and computer ethics* (2008): 69-101.
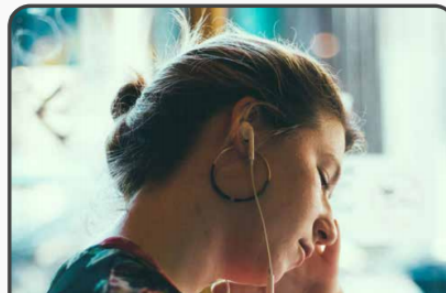
Indirect Stakeholders

Figure 2: Plasma Display Technology Studies

Friedman, Batya, Peter H. Kahn, and Alan Borning. "Value sensitive design and information systems."
The handbook of information and computer ethics (2008): 69-101.
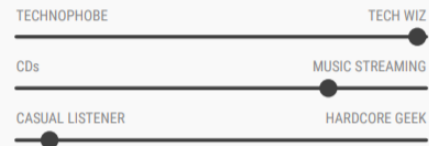
User Personas

Rebecca
Casual audiophile

Age	26
Occupation	Frontend developer
Education	Bachelor degree
Marital status	Single
Location	Mountain View

Online locations	Work and mobile
Computer(s)	iPhone and MacBook Pro
Internet usage	8-9 hours

TECHNOPHOBE ——————— TECH WIZ

CDs ——————— MUSIC STREAMING

CASUAL LISTENER ——————— HARDCORE GEEK

Music is essential to Rebecca's life. She is listening to tunes almost every second of her life, particularly while working.

Obstacles Rebecca faces:
- Too busy to explore new music artists she might like
- Streaming music consumes a lot of data

How will Rebecca interact with Spotify?

Questions Rebecca will ask:
- How do I keep updated on new releases by artists I follow?
- How do I learn of new artists I haven't heard of?
- Can I listen to music in a data-efficient manner?
- How can I listen on both my MacBook and my iPhone?

Who influences Rebecca?

Ads • Friends • Her own tastes • Music charts • Co-workers
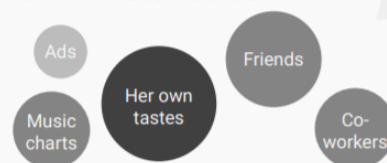
Rebecca's situation

Goals, motivations:
- Listen to great music to keep her productive at work
- Relax and unwind at the end of the day
- Superior music quality for full enjoyment of tracks
- Expand the circle of music artists she listens to

Key words

music, jazz, r&b, pop, artists, new releases, top charts, background music

Rebecca's story

Music is a big part of my life; I like to think that I always have a "background music" running in each scene of my life. I love working while listening to music; somehow, it gives me a lot of focus on my task.

I regularly talk to my co-workers about music and singers — that's what we like to talk about over lunch. We're constantly looking for new artists to inspire us and to expand our music library, but lately it seems a little tough to do that. Everything seems to have a "filter bubble" effect, and we keep listening to the same genres and artists.

I really enjoy finding new artists that match my subjective taste, and most of the times I get those from my close friends. I wish there were a way to find more music and artists without having to rely on the serendipity of life!

Example of a persona that shows the six main elements you should include. Name, age, gender, tag line, experience and skills are placed on the left-hand side. The middle column focuses on the context to indicate how they would interact with a product or service. Finally, on the right-hand side some goals and concerns are shared, as well as a short scenario to indicate the persona's attitude.

https://www.interaction-design.org/literature/topics/personas
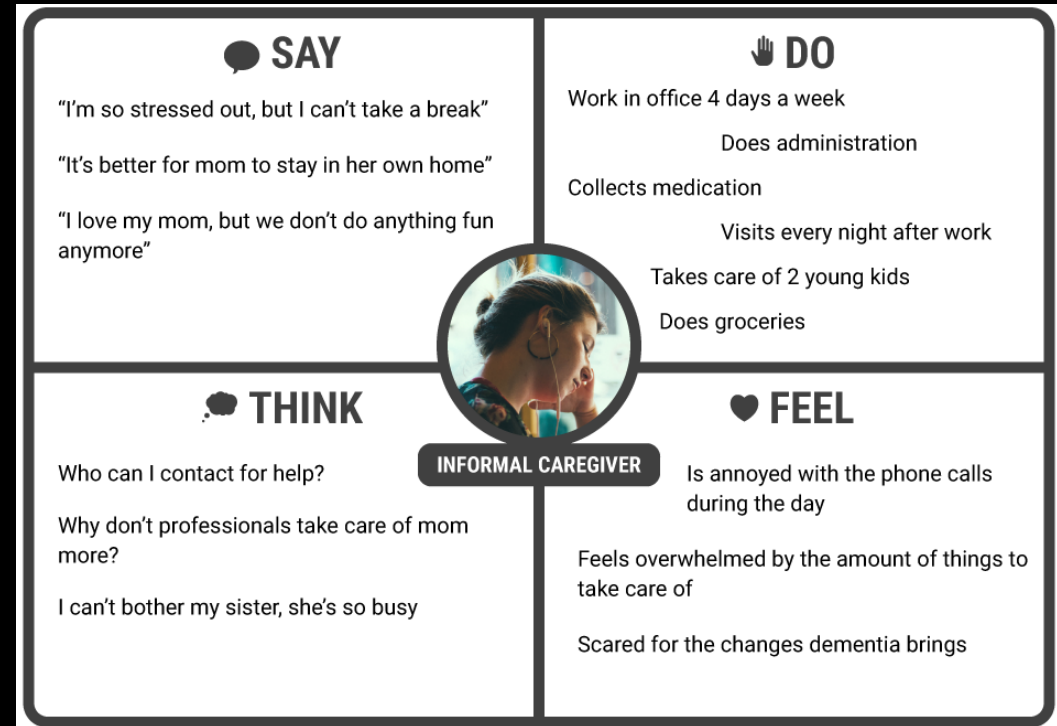
User Personas

- Discovered during the requirements discovery and user story development process

  As a <type of user>, I want <some goal> so that <some reason>.

- Represent *real* behavior patterns, attitudes, skillset, motivations and goals within the product's domain.

- Design for primary personas

User Personas

Empathy

SAY

"I'm so stressed out, but I can't take a break"

"It's better for mom to stay in her own home"

"I love my mom, but we don't do anything fun anymore"

DO

Work in office 4 days a week

Does administration

Collects medication

Visits every night after work

Takes care of 2 young kids

Does groceries

THINK

Who can I contact for help?

Why don't professionals take care of mom more?

I can't bother my sister, she's so busy

INFORMAL CAREGIVER

FEEL

Is annoyed with the phone calls during the day

Feels overwhelmed by the amount of things to take care of

Scared for the changes dementia brings

https://www.interaction-design.org/literature/topics/personas

"On a project for servicing coin-operated vending machines, I introduced Dolly the Serviceperson and Ralph the Warehouse Supervisor. We wrote scenarios for them and they became part of the project team—virtually."

Software Requirements by Karl Eugene Wiegers, Joy Beatty, Microsoft Press, 2013

Requirement Elicitation

A collaborative and analytical process

Interviews

Workshops

Focus Groups

Observations

Questionnaires

System Interface Analysis

User Interface Analysis

Document Analysis

Traps

- Not requirements (during the interactive sessions)

  "Then I select the state where I want to send the package from a drop-down list."

- Assumed, Implied Requirements

- Missing requirements

Acceptance Test

As a consumer, I am always aware of my current energy costs.

As a consumer, I always see current energy pricing reflected on my portal and on-premise devices so that I know that my energy usage costs are accurate and reflect any utility pricing changes.

- Verify the current pricing is always used and the calculated numbers are displayed correctly on the portal and each on-premise device (see attachment for formats).
- Verify the pricing and the calculated numbers are updated correctly when the price changes.
- Verify the "current price" field itself is updated according to the scheduled time.
- Verify the info/error messages when there is a fault in the pricing (see approved error messages attached).

Intersection of User needs and AI strength

As a <type of user>, I want <some goal> so that <some reason>.

Recommending different content to different users?

Prediction of future events?

Recognize patterns in images?

Understand natural language?

……

Maintaining predictability?

Provide static information?

Complete transparency?

……

Dataset Specification and Documentation

- Why is the dataset created? (e.g., is there a specific intended task gap that needed to be filled? Who fund the creation of the dataset?)

- What preprocessing/cleaning is done? (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances)

- If it relates to people, are they told what the dataset would be used for and did they consent? If so, how? Were they provided with any mechanism to revoke their consent in the future or for certain uses?

- Will the dataset be updated? How often, by whom?

Datasheet

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets."

Data Collection Process

How was the data collected? (e.g., hardware apparatus/sensor, manual human curation, software program, software interface/API; how were these constructs/measures/methods validated?)

Who was involved in the data collection process? (e.g., students, crowdworkers) How were they compensated? (e.g., how much were crowdworkers paid?)

Over what time-frame was the data collected? Does the collection time-frame match the creation time-frame?

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part of speech tags; model-based guesses for age or language)? If the latter two, were they validated/verified and if so how?

Does the dataset contain all possible instances? Or is it, for instance, a sample (not necessarily random) from a larger set of instances?

If the dataset is a sample, then what is the population? What was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? Is the sample representative of the larger set (e.g., geographic coverage)? If not, why not (e.g., to cover a more diverse range of instances)? How does this affect possible uses?

Is there information missing from the dataset and why? (this does not include intentionally dropped instances; it might include, e.g., redacted text, withheld documents) Is this data missing because it was unavailable?

Are there any known errors, sources of noise, or redundancies in the data?

Any other comments?

Motivation for Dataset Creation

Why was the dataset created? (e.g., were there specific tasks in mind, or a specific gap that needed to be filled?)

What (other) tasks could the dataset be used for? Are there obvious tasks for which it should *not* be used?

Has the dataset been used for any tasks already? If so, where are the results so others can compare (e.g., links to published papers)?

Who funded the creation of the dataset? If there is an associated grant, provide the grant number.

Any other comments?

Dataset Composition

What are the instances? (that is, examples; e.g., documents, images, people, countries) Are there multiple types of instances? (e.g., movies, users, ratings; people, interactions between them; nodes, edges)

Are relationships between instances made explicit in the data (e.g., social network links, user/movie ratings, etc.)?

How many instances of each type are there?

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)? Features/attributes? Is there a label/target associated with instances? If the instances are related to people, are subpopulations identified (e.g., by age, gender, etc.) and what is their distribution?

Is everything included or does the data rely on external resources? (e.g., websites, tweets, datasets) If external resources, a) are there guarantees that they will exist, and remain constant, over time; b) is there an official archival version. Are there licenses, fees or rights associated with *any* of the data?

Are there recommended data splits or evaluation measures? (e.g., training, development, testing; accuracy/AUC)

What experiments were initially run on this dataset? Have a summary of those results and, if available, provide the link to a paper with more information here.

Any other comments?

Data Preprocessing

What preprocessing/cleaning was done? (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values, etc.)

Was the "raw" data saved in addition to the preprocessed/cleaned data? (e.g., to support unanticipated future uses)

Is the preprocessing software available?

Does this dataset collection/processing procedure achieve the motivation for creating the dataset stated in the first section of this datasheet?

Any other comments?

Dataset Distribution

How is the dataset distributed? (e.g., website, API, etc.; does the data have a DOI; is it archived redundantly?)

When will the dataset be released/first distributed? (Is there a canonical paper/reference for this dataset?)

What license (if any) is it distributed under? Are there any copyrights on the data?

Are there any fees or access/export restrictions?

Any other comments?

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset? How does one contact the owner/curator/manager of the dataset (e.g. email address, or other contact info)?

Will the dataset be updated? How often and by whom? How will updates/revisions be documented and communicated (e.g., mailing list, GitHub)? Is there an erratum?

If the dataset becomes obsolete how will this be communicated?

Is there a repository to link to any/all papers/systems that use this dataset?

If others want to extend/augment/build on this dataset, is there a mechanism for them to do so? If so, is there a process for tracking/assessing the quality of those contributions. What is the process for communicating/distributing these contributions to users?

Any other comments?

Legal & Ethical Considerations

If the dataset relates to people (e.g., their attributes) or was generated by people, were they informed about the data collection? (e.g., datasets that collect writing, photos, interactions, transactions, etc.)

If it relates to other ethically protected subjects, have appropriate obligations been met? (e.g., medical data might include information collected from animals)

If it relates to people, were there any ethical review applications/reviews/approvals? (e.g. Institutional Review Board applications)

If it relates to people, were they told what the dataset would be used for and did they consent? What community norms exist for data collected from human communications? If consent was obtained, how? Were the people provided with any mechanism to revoke their consent in the future or for certain uses?

If it relates to people, could this dataset expose people to harm or legal action? (e.g., financial social or otherwise) What was done to mitigate or reduce the potential for harm?

If it relates to people, does it unfairly advantage or disadvantage a particular social group? In what ways? How was this mitigated?

If it relates to people, were they provided with privacy guarantees? If so, what guarantees and how are these ensured?

Does the dataset comply with the EU General Data Protection Regulation (GDPR)? Does it comply with any other standards, such as the US Equal Employment Opportunity Act?

Does the dataset contain information that might be considered sensitive or confidential? (e.g., personally identifying information)

Does the dataset contain information that might be considered inappropriate or offensive?
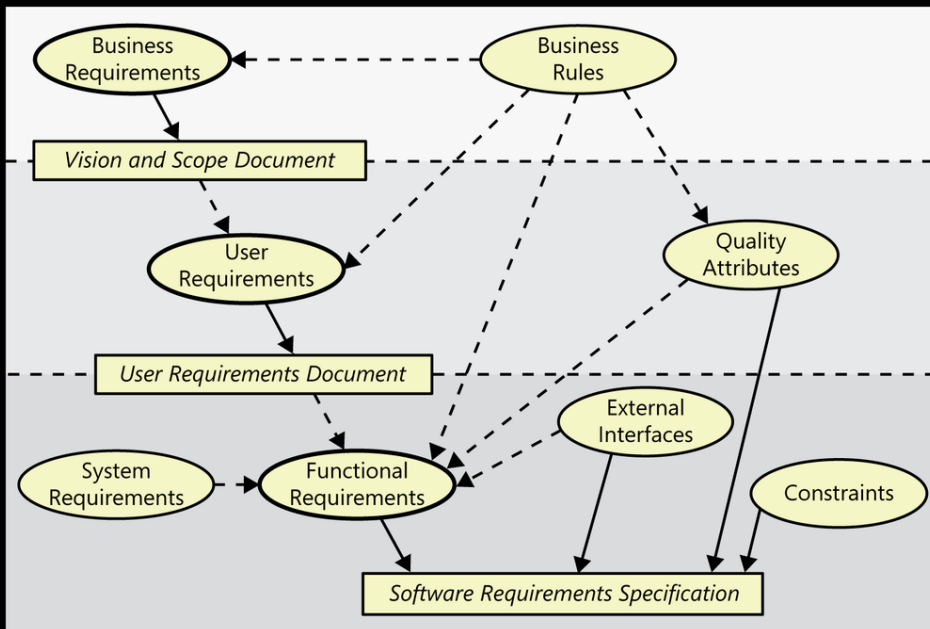
Specification for ML Components
Formal methods perspectives

- System level specification - standard specification languages

- Input-output robustness - Local and Global resilient

- Input-output relations - pre-condition, post condition

- Semantic Invariance - Domain-specific, e.g. scale invariant

- Distribution Assumptions - statistical modelling, e.g. Probabilistic Programming

- Fairness - e.g. individual fairness, demographic parity, etc still evolving area

- ……

Seshia, Sanjit A., et al. "Formal specification for deep neural networks." International Symposium on Automated Technology for Verification and Analysis. Springer, Cham, 2018.

Challenges
Reasoning the relations between different level of requirements

Rahimi, Mona, Jin LC Guo, Sahar Kokaly, and Marsha Chechik.
"Toward Requirements Specification for Machine-Learned Components."
In *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pp. 241-244. IEEE, 2019.

Dreossi, Tommaso, Alexandre Donzé, and Sanjit A. Seshia.
"Compositional falsification of cyber-physical systems with machine learning components."
Journal of Automated Reasoning 63.4 (2019): 1031-1053.