Table 1: List of failure mode as a function of the driver and failure mode-I

Driver	Family	Definition	Example
1	1	Failure to detect different body traits	Hand image recognition is strongly dependent on the hand positioning
1	10	Failure to detect different race traits	Image recognition is strongly dependent on human traits related to fairness consideration
1	1	Failure to detect disruptive traits	Detection failed by devices or traits (e.g. tattoos) that alter the recognition process
1	1	Failure for quantity	Image failure to detect by Lack/poor reverse / flipped image
1	1	Failure for quality	Data used for the training process show lower quality than the used for analyses
1	1	Failure for timeframe representability	The time frames used for training do not match the timeframes of analyses
1	1	Timing gap	Distance between data points does not help to represent phenomena
1	1	Timing	The algorithmic event happens too late or too early, or the control action mistimed
1	1	Lack/poor Functionality	The algorithm provides no output or control action not provided when expected
1	10	Unrecognized bias	Lack/poor recognition or identification of bias from data sources
1	9	Lack/poor data protocols	Lack/poor protocols for data ownership and data responsibilities
1	9	Lack/poor data usability	Data is not related or relevant for the problem to be solveds
1	9	lack/poor data format consistency	Data is not related or relevant for the problem to be solved
1	9	lack/poor data integrity	Data describe altered, unreal, or inconsistent trends in the information supplied.
1	9	Lack/poor temporal data consistency	Data is supplied sporadically
1	9	Lack/poor user responsibilities data	Error applying Data or its processing requirements throughout interphases and designation of responsibilities from the user part
1	9	Data management and responsibility	Poor Data governance from external sources that are dependent on supplied information from/to the AI
1	9	Lack/poor protocols for data validation	No protocols or poor application of them from data validation supplied to the system
1	9	Lack/poor protocols for data curation	No protocols or poor application of them from data curation supplied to the system
1	9	Lack/poor protocols for data tagging	Lack/poor methods to track data modifications, if allowed, by tagging and users identification
1	9	Lack/poor protocols for data tagging	Lack/poor methods to track data modifications, if allowed, by tagging and users identification
1	4	Lack/poor internal data or algorithmic responsibility	Poor or lack of designation of responsibilities for internal data sources management, quality, veracity, and quantity.
1	4	Lack/poor external data or algorithmic responsibility	Poor or lack of designation of responsibilities for external data sources management, quality, veracity, and quantity.
1	10	sensitive information indirect disclosure	Information released in anonymised form might lead to disclosure) of personal data (if linked through different algorithms or metadata. usage
1	10	Lack of Data bias reduction	No approach or algorithm implemented for bias/unfairness reduction within data
1	3	lack or poor data collecting labelling	Inadequate or nonexistent data labelling (and or metadata) to track the origin (and times) of data collection.
1	3	lack or poor data processing labelling	Inadequate or nonexistent data labelling processed by the AI.
2	1	Timing gap data usage	Lag or mismatch on timeframes between information capture (sensoring) and use of it
2	1	Hardware requirement	The hardware or system in which the algorithm components are kept are insufficient for the requirements
2	1	Interface Failure	Software failure due to failure of hardware interfaces such as power supply
2	8,9	Information physical disclosure violation	Accidental loss of electronic equipment by personnel that contain algorithmic and/or database
2	10	Lack of Data separation	All information is kept on same physical/cloud component with same level of accessability indepennt of its sensitivity
3	1	Sequence	Algorithmic event occurs in the wrong order or control action with incomplete sequence concept error
3	1	False positive detection from alarm/action	The algorithm detects an error when there is no error or control action provided when not expected
3	1	False-negative detection from alarm/action	The algorithm does not detect an error when there is an error or control action provided when expected
3	1	Fault logic and Ranges	Concept error where the software or control actions contain incomplete or overlapping logic
3	1	Incorrect computation from recognised input	Incorrectly computes based on some or all inputs or control actions. The potential source of error is identified
3	1	Incorrect computation from unrecognized sources	The software computes incorrectly. The potential source of error is NOT identified.
3	1	Memory Management	The algorithm performs actions that make the system run out of memory
3	5	Lack/poor use or misuse of societal metrics	No use of features for tracking and reporting on social trends or impacts
3	6	Lack/poor use or misuse of env. metrics	No use of features for tracking and reporting on env. trends or impacts
3	10	Incomplete data sets	unrepresented clusters or groups by an uneven data
3	10	Lack/poor bias elimination	Lack/poor methods or approaches to eliminate biased data from data sources known to contain them
3	3	Lack/poor Transparency Algorithmic approach	There is no explainability approach or is not suitable for their user interpretation/understanding
3	1	Common corruption	The system is not able to handle common corruption and perturbations such as tilting, zooming, or noisy images
3	3	Inappropriate algorithm used for explainability	Algorithm used for is not the most suitable one for the data type managed by the AI (i.e. natural language, image, tabulated).
3	10	Lack of Algorithm bias reduction	No approach or algorithm implemented for bias/unfairness reduction within data
3	10	Lack/poor bias reduction approaches for AI	No approach, safeguards or algorithm implemented for bias/unfairness actions or estimations performed by the AI
3	1	Common corruption	The system is not able to handle common corruption and perturbations such as tilting, zooming, or noisy images
4	1	Improper Functionality	The programmed control system software performs an unexpected action as defined by the user
4	1	Lack/poor algorithmic corrective actions	Lack/poor identification and action associated with protecting the algorithmic robustness e.g. maintenance
4	5	Lack/poor social metrics	No features for tracking and reporting on social trends or impacts
4	6	Lack/poor environmental metrics	No features for tracking and reporting on environmental trends or impacts
4	9	Lack/poor use or misuse of societal metrics	No organizational policy for the protection of property (prevent steeling of technical resources)
4	10	Human Rights communication and AI ethics	Lack/poor understanding/violation of required topics by developers (e.g. human rights)
4	5	General communication problems	Lack/poor processes for resolving grievances from AI
4	5	Lack/poor regulation compliance	Conditions of work with the AI do not comply with local or regional law/requirements
4	5	Lack/poor fair operating practices	Error, no driver, or no methodologies to apply corrective actions to fairness
4	5	Lack/poor fair operating practices	Error, no driver, or no methodologies to apply corrective actions to fairness
4	2	Lack/poor security and safety corrective actions	Lack/poor identification and action associated with protecting the algorithmic robustness and users
4	9	Lack/poor governance corrective actions	Lack/poor identification and action associated with securing data governance
4	4	Lack of accountability corrective actions	1 Lack of identification and action associated with securing data accountability for data and algorithms
4	3	Lack of Transparency in corrective actions	Lack of identification and action associated with securing system transparency in algorithms

Drivers: 1 Data driver, 2 Physical Driver, 3, Algorithm, 4 Internal Social, 5 User and System Interphase.
Family: (1) Robustness, (2) Safety, (3) Transparency, (4) Accountability, (5) Societal well-being, (6) Environmental well-being, (7) Human agency and oversight, (8) Privacy, (9) Data Governance, (10) Unbias, and (11) Users Values.

Table 2: List of failure mode as a function of the driver and failure mode - II

Driver	Family	Definition	Example
4	5	Lack of (5) corrective actions	Lack of identification and action associated with securing societal wellbeing for data and algorithms
4	7	Lack of (7) corrective actions	Lack of identification and action associated with Human Agency and Oversight
4	10	Lack of bias corrective actions	Lack of identification and action associated with a bias from data, developers, and algorithms
4	11	Lack of User Values corrective actions	Lack of identification and action associated with users' values and its trends for data, developers, and algorithms
4	8	Lack of privacy corrective actions	Lack of identification and action associated with data privacy
5	1	Improper software use	Requirements set by users are not achievable by the algorithm or its scope set for training
5	9	Internal unauthorized action	Interphases not responding or not connecting by unauthorized users or actions (e.g. rebooting actions)
5	9	Lack/poor user responsibilities actions	Error actionability throughout interphases and designation of responsibilities from the user part
5	9	Lack/poor accessibility protocols	Lack/poor user recorder info protocol for securing user identification
5	9	Lack/poor accessibility protocols	Lack/poor protocol for securing user access
5	9	Over accessibility	Lack/poor control of the user and developers' access to restrictive information, source code, and algorithmic parameters
5	3	Transparency Acceptance	Users do not accept explanations or outcomes produced by the algorithm
5	1,2,8	Perturbation attack	The attacker modifies the query to get an appropriate response
5	1,2,8	Poisoning Attack	Attacker contaminates the training phase of ML systems to get the intended result
5	1,2,8	Model Inversion	Attacker recovers the secret features used in the model
5	1,2,8	Membership Inference	Attacker infer if the given data record was part of the model's training data set
5	8	Model Stealing	The attacker can recover the model by constructing careful queries
5	1,2,8	Reprogramming ML system	Repurpose the ML system to perform a non-programmed activity
5	1,2,8	Adversarial Example in Physical Domain	Attacker brings adversarial examples into the physical domain to subvert ML system
5	1,2,8	Malicious Recovering Training Data	Malicious ML providers can query the model used by the customer and recover the customer's training data
5	1,2,8	Attacking the ML Supply Chain	Attacker compromises the ML model as it is being downloaded for use
5	1,2,8	Backdoor ML	Malicious ML provider backdoors algorithm that does not work unless triggered
5	1,2,8	Exploit Software Dependencies	The attacker uses traditional software exploits to confuse ML systems
5	1,2,8	Reward Hacking	Reinforcement learning systems act in unintended ways because of a mismatch between stated reward and true rewards.
5	1,2,8	Side Effects	System disrupts the environment as it tires of attaining its goal
5	1,2,8	Natural adversarial examples	Without attacker perturbations, the ML system fails to owe to hard harmful mining
5	1	Incomplete testing or training	The Ml systems are not tasted or trained in realistic conditions that it is meant to operate
5	1, ... ,11	User protocols or definitions missuses	Violation of algorithms or methods by users intentionally or unintentionally causes failure by *i* (i within the family from 1 - 11).
5	3	lack human direct communication	the system only allows communications with the AI when human-to-human interactions are possible.
5	7	unrecognized AI	no clarifications that processes or actions are led by AI (e.g. chatbot or optimized results in Human-on-the-loop approaches)
4,5	8,9	Information internal disclosure violation	Inappropriate disclosure of personal data internally within system developers by lack of appropriate controls
4,5	10	sensitive information accesible	Vulnerable or sensitive data kept is accessible or inappropriate disclosure
4,5	8,9	Data disclosure by a change in AI conditions	Personal data being used in a manner not anticipated by data subjects due to change on AI functionalities or change on internal policies.
4,5	3,8,9	failure to explain effectively data usage	Information is not used as expected by users (8), algorithms (3), or developers (9).
1,5	5	automated intrusive data usage perception	Personal data being used for automated decision-making may be seen as excessively intrusive. The perception (or historical misuse) causes system rejection
1,5	11	values violation perception	Users value is perceived as violated by the algorithmic results, data or the processes involved in
1,5	8	privacy disruption by data merging	Merging of datasets may result in more individual information than anticipated by the users or developers.
1,5	8	privacy disruption by data merging	Merging of datasets may result in more individual information than anticipated by the users or developers.
1,5	8	individual recognition by data merging	individuals can be identified by data merging
1,5	8,9,11	inappropriate data sensing approach perception	visual or audio recordings may be perceived as unacceptably intrusive given the system and environment condition (e.g. google glass).
1,5	8,9,11	Anonymously blockage	Data collection containing identifiers may prevent users from using a service anonymously when type and level of access is not relevant
1,4	4,9	Data Mismanagement policies	Data may be kept longer than required in the absence of appropriate policies for its elimination.
1,4	4	Data Mismanagement processes	processes to remove records are inexistent or not executed correctly
1,5	4, 9	Unneeded data management policies	Data unnecessary for the sytem are collected given poor or lack of policies in place
1,5	4, 9	Unneeded data managing processes	processes for extracting/eliminating useless information are not applied correctly
1,5	4, 9	Lack of regional data transfer policies	Unexistance of regional/ local data management policies regarding data transfer.
1,5	4, 9	Violation of policies and regional regulations regarding data transfer and usage	Transfer of private information collected in one region and saved in another. This includes the physical transfer or cloud/on-line data transfer.
1,4,5	4,9	unintended data Duplication policies	Data is duplicated and saved in different locations, giving poor/lack of policies of management and use of historical records. This considers data processing and accumulated original and modified data on new files
1,4,5	4,9	unintended data Duplication processes	Data is duplicated and saved in different locations, giving violation of management approaches of historical records.
3,4,5	3,4,7	system barrier of understanding	Not understanding of system functionalities or outputs are given language barrier
4,5	3,4,7	system's development/maintenance documentation policies	poor or nonexistent system documentation policies that can be used to explain system errors.
4,5	3,4,7	system's development/maintenance documentation processes	violation of documentation processes for AI development or usage, if needed.
2,5	7	unmatch human-AI response time	Inadequate available time response for human-AI interactions in human-in-the-loop, human-on-the-loop, or human-in-control
2,5	7	Incorrect human interaction dynamic type	System dynamics are inadequate for human intervention even when different approaches have been defined for human intervention
1,2,3,5	1, 4	Incorrect Failing Detecting Approaches	There are not enough approaches to detect failing conditions and thus, secure action devices. These include alarm systems, system brokers, and others.

Drivers: 1 Data driver, 2 Physical Driver, 3, Algorithm, 4 Internal Social, 5 User and System Interphase.
Family: (1) Robustness, (2) Safety, (3) Transparency, (4) Accountability, (5) Societal well-being, (6) Environmental well-being, (7) Human agency and oversight, (8) Privacy, (9) Data Governance, (10) Unbias, and (11) Users Values.