Alexandra Karakas
PhD Student, Eötvös Loránd University

Malfunction notions in AI and design

Although, at first sight, malfunction – especially from a user's perspective - seems to be a simple concept, but in reality, in both philosophy of technology and engineering failure as such is a wide-ranging phenomenon. Malfunction can refer to several different anomalous scenarios, including poor design, manufacturing mistakes, unexpected features, and unplanned user reaction. What is common in all cases is that malfunction indicates an artefact's inability to perform its intended function. I use the term function as a concept that refers to the activity/activities what the artefacts are both designed for and treated by the user, and at the same time describes the artefact's means-to-end nature, and thus it gives an explanation of the object.

Broadly speaking, malfunction in AI/computer engineering can be divided into two types: (1) hardware errors and (2) software malfunction. Regarding (2) AI engineers tend to use the word *bug* when any kind of error occurs in a software for instance, while in the case of (1), hardware, and design, failures are described usually as malfunction. As there is no taxonomy for malfunction yet, the boundaries are blurred between the different understandings of malfunction in technology. In my talk I focus on the differences between malfunction notions in physical objects and in contrasting categories such as software design and AI. Is there a fundamental distinction between the notion of failure in physical artefacts and software engineering, or the relation between them is more dialectical? The talk explores the basic differences between (1) artefact malfunction and (2) AI failure.