# Mushroom Classification

**Mushrooms are the visible fruiting bodies of fungi, serving as reproductive structures that release spores from gills or pores for dispersal. Most belong to the Basidiomycota or Ascomycota phyla and exhibit diverse morphologies—caps, gills, stipes, and sometimes volvas—that are key to species-level identification.**

Accurate identification is vital because morphology alone cannot distinguish edible from poisonous species consistently. Some deadly fungi such as *Amanita phalloides* (“death cap”) closely resemble edible forms and contain heat-stable amatoxin compounds that cause severe liver failure and can be fatal even in cooked specimens. Misidentification remains the leading cause of mushroom poisoning. Biological expertise in gill attachment, spore print colour, odour, bruising reactions, and habitat, alongside microscopic or genetic confirmation, dramatically enhances both **safety** and **understanding of fungal diversity.**



Figure 1: Amanita Phalloides - the "Death Cap" - is a highly toxic mushroom in the Amanita family.

This dataset features 61069 records of mushrooms, with 23 features identified for each record. The full set of data can be found in [mushroom\_observations.csv](mushrom_observations.csv). Also included in this dataset is a list of the characteristics of the 173 individual species of mushroom represented in the dataset, found in <mushroom_characteristics.csv>.

The dataset has good potential to build a machine learning model which links visual characteristics of mushrooms with their edibility.

**Data Description**

The columns in the observations dataset are mostly categorical, with letters representing specific categories. For example the “gill-spacing” category describes the visual characteristics of the gill spacing for a single mushroom, and can take values “c” (close), “d” (distant) or “f” (none). Some categories take continuous numerical values, such as “stem-length” and “cap width”. For a full description of the dataset, see <observations_metadata.txt>.

The characteristics dataset gives the range of potential visual characteristics for the 173 species represented in the observations dataset. This is represented as a range for numerical variables, or a list of potential values for categorical variables. For example, the Fly Agaric mushroom can have a cap diameter between 10 and 20 centimetres, so the record in the “cap-diameter” column reads “[10, 20]”; the cap surface is shiny with grooves, so the “cap-surface” column reads [g, h]. See <characteristics_metadata.txt> for a full breakdown of the data.