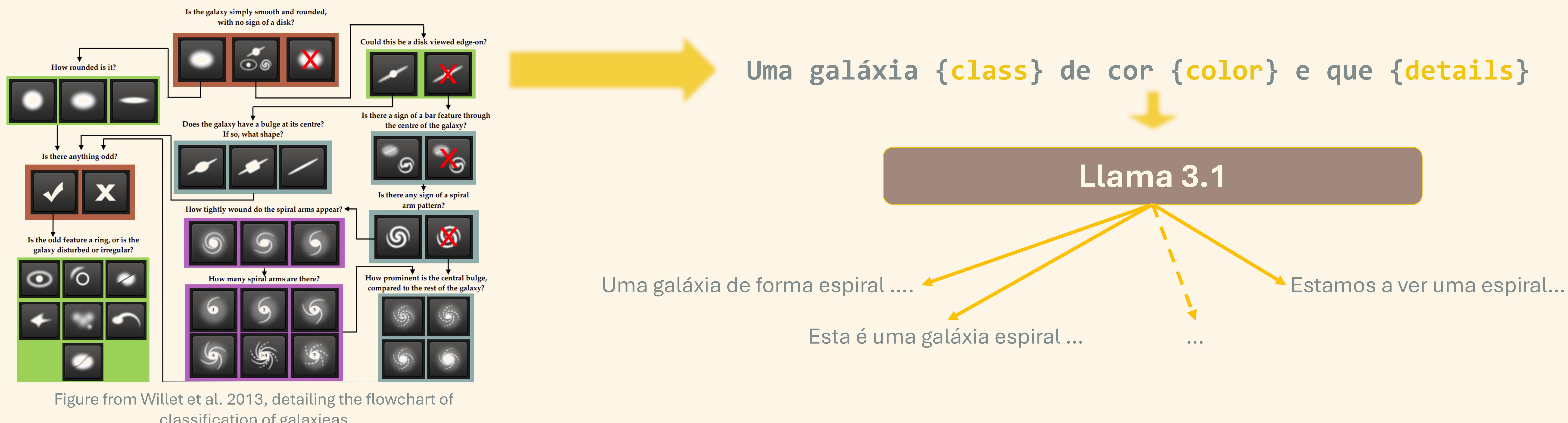


AstroLingua: Empowering Inclusive Astronomy through AI

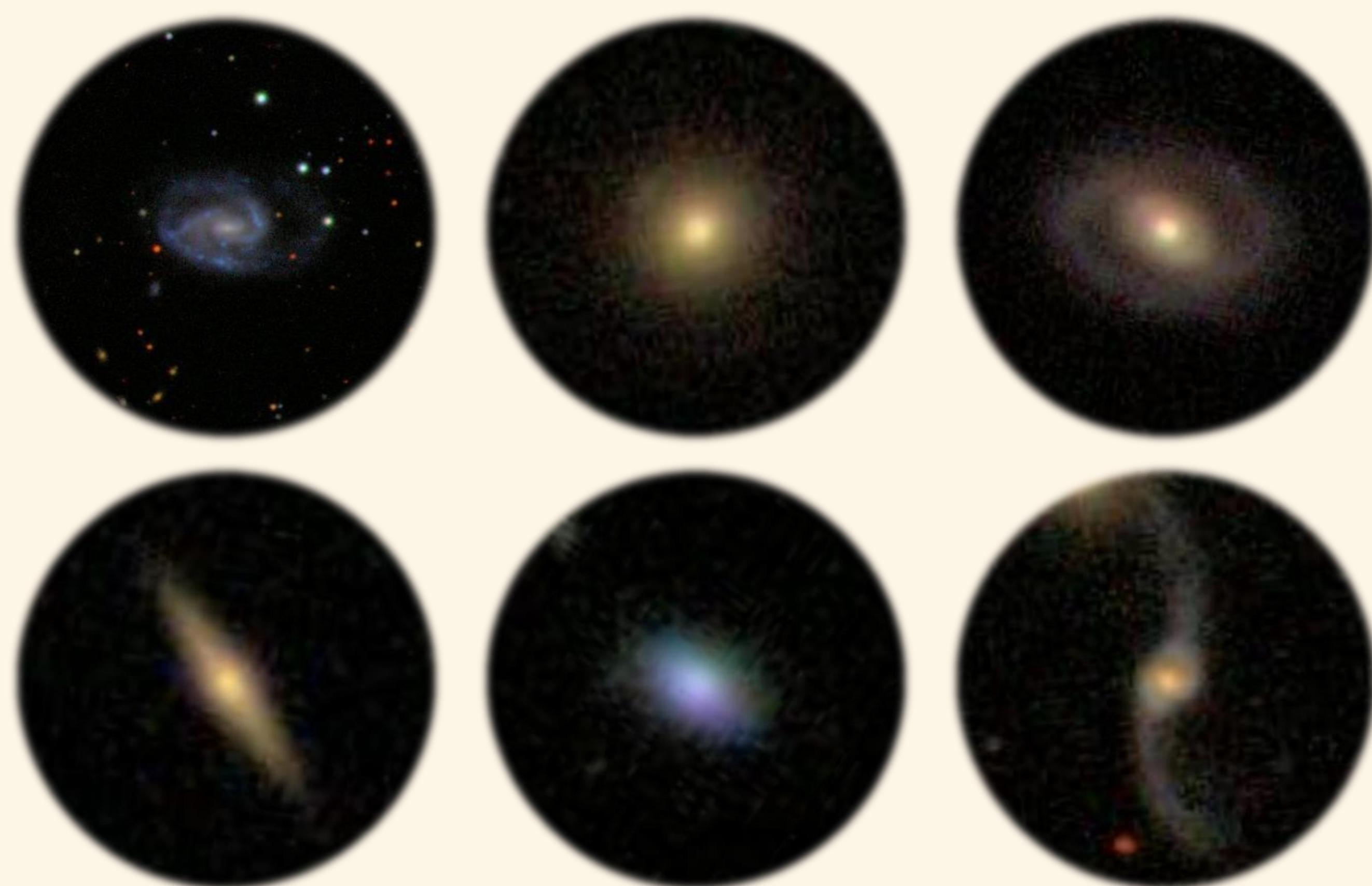
A. Paulino-Afonso (IA-UPorto); Pedro Cunha (IA-UPorto); Bruno Ribeiro (Celfocus) & AstroLingua Team

Images of extragalactic objects have captured the fascination of the public, giving rise to a new generation of citizen science projects. The Galaxy Zoo project, gathering nearly one million classifications worldwide, provides a diverse dataset of galaxies in various shapes and sizes and cosmic times. The **AstroLingua** project aims at developing a comprehensive pipeline that seamlessly integrates state-of-the-art image analysis techniques and natural language processing (NLP) to deliver detailed descriptions of galaxies based on their morphology and environmental characteristics.

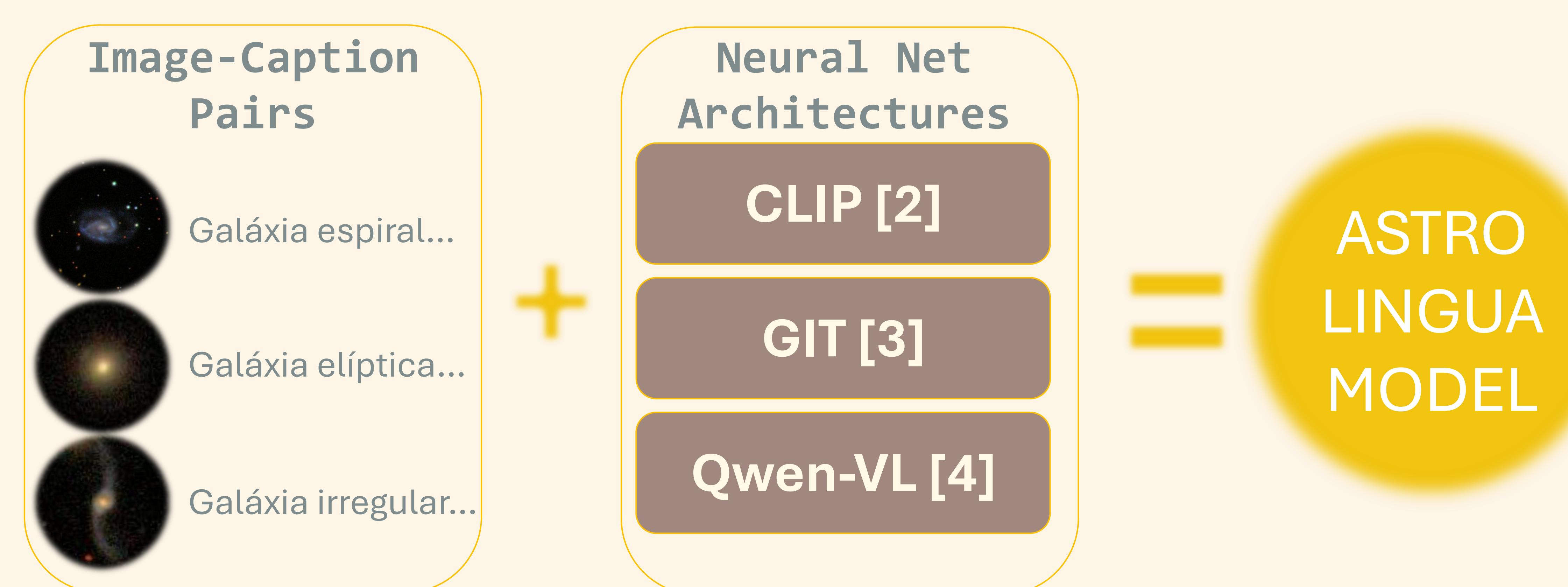
1. Use Galaxy Zoo [1] classifications to generate base template captions.
Then use a generic LLM to create variations for data augmentation



2. Collect Images from Galaxy Zoo SDSS

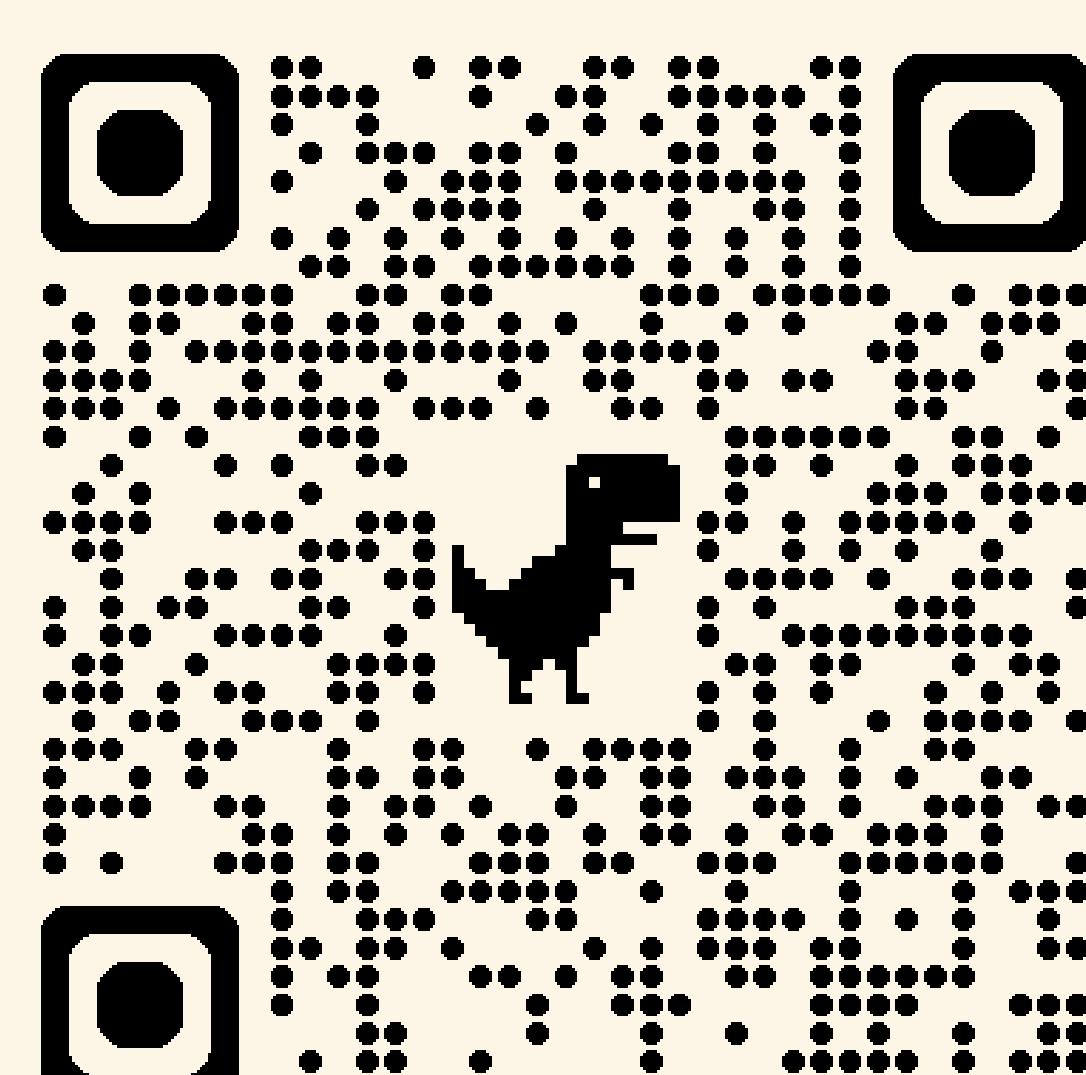


3. Fine-tune Image-to-text model



4. Make the model publicly available for download and through a web application for dissemination

Try it yourself here



The screenshot shows a user interface for generating captions. At the top, there is a button "Select New Random Image". Below it is a large image of a spiral galaxy. To the right of the image, there is a "Caption" section with two parts: "English:" and "Português (Portugal)". The English caption reads: "The central galaxy appears to be a spiral galaxy with a bright, yellowish core and loosely defined spiral arms. The galaxy displays a noticeable dust lane, a darker band of dust that bisects the core. The galaxy's arms appear to be somewhat faint and diffused, suggesting it might be a relatively low-mass galaxy or viewed edge-on. The overall color of the galaxy appears to be a yellowish-white, indicating a relatively old stellar population." The Portuguese caption reads: "A galáxia central parece ser uma galáxia espiral com um núcleo brilhante e amarelo e braços espirais vagamente definidos. A galáxia exibe uma faixa de poeira notável, uma banda de poeira mais escura que divide o núcleo. Os braços da galáxia parecem ser um pouco ténues e difusos, sugerindo que pode ser uma galáxia de massa relativamente baixa ou vista de perfil. A cor geral da galáxia parece ser branca-amarelada, indicando uma população estelar relativamente antiga." At the bottom of the interface, there is a small note: "image: static/images/1074.jpg".