The rapid proliferation of highly realistic speech deepfakes poses a growing threat to information integrity and public trust. Current detection systems are predominantly optimized for English and, due to the scarcity of non-English benchmarks, they generally suffer significant performance degradation when applied to other languages. To address this, we introduce *ITASpoof*, a large-scale dataset comprising over 1 million real and synthetic Italian speech samples. With real samples collected from the "*Corpora e Lessici di Italiano Parlato e Scritto*" corpus, the dataset ensures gender balance while capturing the linguistic and geographical diversity of 15 Italian cities. Synthetic speech is generated using seven state-of-the-art zero-shot Voice-Cloning TTS methods, reflecting the current capabilities of these generative systems. ITASpoof comprises two subsets: *Detection*, designed to evaluate the discrimination between real and synthetic speech, and *Attribution*, which benchmarks the ability to identify the specific generative model used for synthesis. Complementing the dataset, we propose a detection framework following an Out-of-Distribution (OOD) training paradigm. This approach is designed to enhance robustness against unseen generative techniques while improving cross-language generalization. Preliminary experiments support the long-term goal of developing a truly language-agnostic detector, capable of effective detection without language-specific fine-tuning.