Analysis of Log Data to Identify Productive Strategies in Programming Tasks of an International Study

Professor:	*Francesco Mondada*	*ME B3 426*	*francesco.mondada@epfl.ch*
	Engin Bumbacher	*HEP Vaud*	*engin.bumbacher@hepl.ch*
Scientific Assistant Contacts:	*Jérôme Brender*	*ME B3 435*	*jerome.brender@epfl.ch*

Project Type: ***Semester Project (Spring)*** *Section: Data Science, Computer Science, Communication Systems Micro & Robotics (IC, STI)*

Official Start Date: *19.02.2024*

Programming skills are essential to thrive in today's world. The OECD PISA is developing a new platform for teaching and measuring important 21st-century skills, for which we have developed a first application named Karel (see Figure 1) to teach and assess programming skills. We have already conducted an international study with hundreds of pupils using the application, which has shown promising results in helping learners develop these skills. The core of this project is the rich dataset gathered from these young coders. For each pupil, we have logged data on how they approached the programming problems. We want to use this data to develop algorithms that help us identify productive problem-solving strategies. This will assist us in developing systems and teaching approaches to help students learn how to solve programming problems, which will be useful for schools around the world!

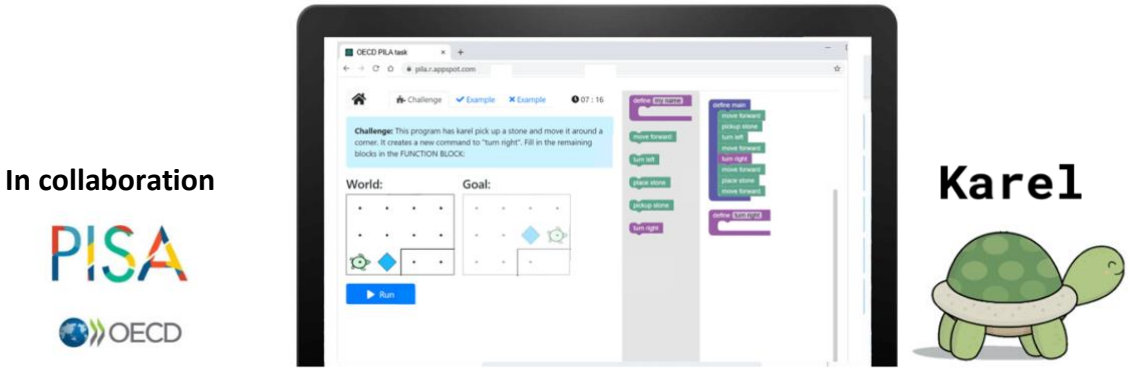

Figure 1: *Example of one task with the current Karel*

Your mission (based on prior work):

- **Data Pre-processing**: Familiarize yourself with the pupil dataset generated from the application interface. Undertake filtering and cleaning procedures to eliminate noise and outliers, ensuring the data's integrity.
- **Data Visualization & Analysis:** Focus on iteratively designing, developing and transforming data into insightful visual narratives. Your goal is to aid educators in recognizing effective problem-solving techniques through representations of pupil progress and code. Utilize programming languages like Python or Java to develop these systems.

This is a multidisciplinary and cutting-edge project that integrates CS and Learning Sciences and has an international scope. You will have the opportunity to work with people who have experience in similar projects for guidance and brainstorming. You will be given freedom in terms of proposing intuitive user interfaces, choosing the libraries, and designing systems. This is an ideal project for students interested in data science and computer science with real impact.

Helpful (but not mandatory) prerequisites for students:

- Good coding skills, ideally in Python.
- Basic knowledge of data analysis (ideally with experience in Applied Data Analysis (ADA))
- A keen interest in education and a creative mindset.
- Proficiency in English; knowledge of French is a plus.