Table 1: **Comparing M3R Processing Pipelines.** NDMG uniquely achieves all of the design desiderata for both diffusion and functional MRI. offers a similar or better range of deployment options compared to cutting edge neuroscience pipelines. Below, modern neuroscience pipelines are compared across a variety of algorithmic and deployment-related characteristics.

		reliable				scalable				portable				open	
pipeline	accurate	TRT	ICC	robust	expedient	b	e	m	c	dc	sc	vdi	turn-key	os	d
NDMG-D	✓	0.98	–	10	O(1 hr)	✓	–	✓	–	✓	✓	–	✓	✓	✓
NDMG-F	✓	0.88	–	18	O(1 hr)	✓	–	✓	–	✓	–	–	✓	✓	✓
CPAC	✓	0.88*	–	18	O(1 hr)	✓	✓	✓	✓	✓	✓	–	–	✓	✓
fmriprep	✓	–	–	?	O(1 day)	✓	✓	–	✓	✓	✓	–	✓	✓	✓
PANDA	✓	–	–	–	O(1 day)	–	–	✓	✓	–	–	✓	?	?	?
mindboggle	✓	–	–	1	?	–	–	–	–	✓	–	–	✓	✓	✓
HCP Pipeline	✓	–	–	✓	–	–	–	–	–	–	–	–	–	✓	✓
NIAK	✓	–	0.46	1	?	–	–	–	–	✓	–	–	–	–	✓

The rows in the above plot were chosen as pipelines that implement either diffusion or functional MRI processing. The top two rows are implemented by this package. The CPAC pipeline provides a fMRI processing pipeline well-suited for a range of processing options, as it enables users to choose optional algorithms of choice. The fMRIPREP pipeline provides cutting-edge fMRI processing for users to leverage the latest processing tools on high-quality datasets. the PANDA pipeline provides introductory exploration for diffusion MRI. The mindboggle pipeline allows users to analyze T1w data to identify anatomical regions suitable for future explorations. The HCP Pipelines provide basic minimal processing and was used on the extensive HCP dataset. The NIAK pipeline provides a a neuroimaging analysis toolkit aimed at the collection of statistics in the pre-processing phase.

The columns of the above plot were chosen to highlight desirable algorithmic and deployment-related characteristics of a processing pipeline. Accuracy measures the distance between the "truth" and an estimate produced by an analysis pipeline. While accuracy is difficult to quantify in a neuroimaging context, we focus on the subjective accuracy of each processing stage by the production of quality assurance (QA) and quality assurance statistics (QAX) throughout the course of the analysis. Reliable refers to whether the pipeline produces a similar result given a similar input. Here, we consider two methods, the test-retest reliability (TRT) and the intraclass correlation coefficient (ICC), which each provide a statistic indicating some comparison of the intra (within)-subject and inter (between)-subject variability. Robustness refers to whether the pipeline has been run and evaluated on a variety of datasets. It is important for pipelines to be simple to deploy on a range of subjects, or scalable to more inputs. We compare scalability across AWS Batch (b), AWS EC2 (e), multi-threading (m), and cluster usage (c). Pipelines are evaluated for their portability to examine the ease of using the pipeline on a variety of systems. The pipelines are compared for several characteristics that reduce the installation requirements for users; namely, whether they provide docker containers (dc), singularity containers (sc), or virtual disk images (vdi). Turn-key methods do not require users to specify parametrs and settings for each stage of processing, meaning users do not have to fine-tune the pipeline to their specific applications. Finally, openness refers to the ease with which users can contribute to the scientific progress. The pipelines are evaluated for open-source code (os), and open-source derivatives (d), which allow users with a computer to contribute to the code base or begin to analyze derivatives immediately.

By developing NDMG according to these algorithmic and computational design principles, and running it on so many datasets, we are able to use it to evaluate subjects, groups, and the collection of groups at an unprecedented scale. It is clear that the NDMG pipeline uniquely satisfies each of the desirable desiderata categories (accuracy, reliability, robustness, expedience, scalability, portability, turn-key, and openness) in some form. The only other pipeline to come close to the NDMG pipeline in terms of the desiderata is the CPAC pipeline. Based on the above evaluation, possible action to take would be to implement an EC2 instance containing the NDMG pipeline. This would simplify the EC2 launch procedure if users wish to have an interactive container running the application and not require technical knowledge of how to deploy docker containers in EC2.