Technical due diligence report for early stage investors

Company: X
From: Thomas Wood, Director: Fast Data Science Ltd

To: X

Date: X

This template is designed for technical due diligence on companies using machine learning. It does not include the financial or legal questions that form the main part of a due diligence investigation, nor does it address the marketability of a product.

Executive summary

Architecture and code

Is the software documented?	
Code maintainability	
Unit tests	
Version control and ticketing	
Look in Git: who has committed and how frequently?	
Is it possible to understand the code with no introduction?	
Legacy or obsolescence?	
Remarks	

Machine learning model

Is the model accurate enough for its intended purpose? Does it work?	
Are the algorithms documented and explainable?	
Is it possible for the user to verify the model results?	
Is a sample test set available to verify model performance? Is this completely separate from any training data?	
Is the machine learning model under version control?	
How will the model be retrained?	
What KPIs will be used to score the model?	
Will there be a closed loop of data coming in for retraining after deployment?	
How well does the model adapt to changes in its environment after deployment?	
If so, how will future retrained models be quality controlled?	
Is AI bias likely to be an issue?	
Accuracy, precision/recall, ROC, AUC etc?	

Does the model output a degree of confidence in its predictions?	
Does the model require fallback to human intervention? How is this fallback decided?	
Remarks	

Technology used

Is the system built on mainstream AI technologies which are expected to be well-supported in future?	
Are the solutions documented (separately from code documentation), so that if the company was acquired or personnel leave, it can still be operated?	
Remarks	

Data processing

Is the data processed in the cloud or on premise?	
Remarks	

Data

Has enough data already been collected?	
How can data quality be assured?	
Cold start problem?	
Is data collected so far specific to one customer or location and unlikely to generalise to new situations?	
Is there a bottleneck in data gathering if humans are needed to label every instance?	
Is there a risk of cross-contamination, e.g. data of one customer being used on another customer?	
Has all data been appropriately acquired (not been scraped or otherwise obtained without permission)?	
Has any data been gathered via user devices such as mobiles, and has permission been granted for this?	
Remarks	

Scalability and ML Ops

Will the services need to be separated across servers?	
Will sessions be stored?	
How will costs scale as the operation scales? (Pay special attention to cloud computing costs)	
What is the availability of the system as a percentage?	

Is there a single point of failure?	
Is there a disaster recovery plan and has it been tested?	
How can bugs be reported and how can tech issues be resolved in future?	
Remarks	

Intellectual property

Is the technology patented?	
Is the technology easy to copy?	
Is it hard to obtain data to train this kind of model?	
Are there any limitations in terms of licenses of software or patents that are being used?	
Does the technology depend on third-party or open-source code, model architecture, dataset, or model weights?	
Was transfer learning used to develop the model, or was it trained from scratch?	
Any vendor lock-in?	
Does the company own all inventions of its employees? (E.g. if an employee is also active in a university, the university may own IP)	
If any open source component is used, does this cause a cascade of obligations to make the derived product also open source?	
Remarks	

People

Who are the most important members of the team?	
Are their credentials and experience valid as claimed?	
What are their engagements? Full time, part time?	
What are their responsibilities?	
Who is engaged with this company as their primary activity?	
Is anybody crucial to the organisation? Any bus factor of 1?	
Any employee with e.g. exceptional talent for writing code?	
Any skills gaps?	
Any redundancies? Do team's skills complement each other or do they overlap (e.g. 100% PhD mathematicians or 100% developer background)?	
Remarks	

Product support

How will tech issues be resolved in the future?	
Remarks	

Regulation

Does it need regulation of any kind?	
Are we dealing with personal data?	
Do we need to consider environmental regulations? (if applicable)	
Is it licensed as a medical device? (if applicable)	
What needs to be done to be able to sell it as a **non-medical** device? (if applicable)	
What additional approval would be needed to be able to sell it as a **medical** device? (if applicable)	
Are there any applicable AI regulations, such as ISO/IEC JTC 1/SC 42, and EU regulations (the AI Act, the Digital Services Act, the Digital Markets Act, the Data Governance Act, EU Data Act), UK regulations (AI Regulation Policy Paper), US regulations (Bill of Rights).	
Remarks	

Ethics

Is there a person in the organisation responsible for data ethics?	
Can the environmental impact of the ML life cycle be measured?	
Are there any measures to reduce the environmental impact?	
Remarks	

Future roadmap of the company

What are the plans for the next few years?	
Is there a well-documented roadmap for new features?	
How many resources are allocated to new development?	
Which key features will be launched within the next 12 months?	
Remarks	

Risks

What technical risks are possible?	
Remarks	

GDPR

Is GDPR/HIPAA/data protection relevant for this venture?	

Is the company a data controller or data processor, and have they registered with the relevant authorities such as the Information Commissioner's Office?	
Does company have a Data Protection Officer?	
Are the staff aware of their data protection obligations?	
Right to be forgotten: would it cause a problem if a user wanted themselves removed from all datasets?	
Risk of data leak? Subjects must be notified if a data leak occurs. Are datasets secure?	
Subject access request: can the company provide an individual with all their data?	
Could sensitive data be reconstructed by reverse engineering a model?	
Is data processing transparent and fair?	
Is consent requested and recorded where needed, and have users consented to storage and processing of their data?	
Is data sent across borders?	
Is there a secure deletion process?	
Remarks	

Recommendations for the legal team

GDPR/HIPAA/data protection questions to investigate?	
Software licences to investigate?	
Patents to investigate?	
Remarks	

Recommendations for the financial team

Anything to look for in accounts?	
Remarks	

Business opportunities

Who are the main competitors, if any?	
Can the software be used for other purposes?	
What are the key technical barriers to entering a new market, such as a new country?	
Remarks	