

AI-Enabled Document Analytics

8 Nov 2023





DISCLAIMER

By acceptance of this Document, the recipient agrees that this document and the information contained therein, and the recipient's involvement in the process described herein will be kept confidential and not disclosed to third parties unless agreed in writing by the author. All information contained in this Document is Confidential Information.

Use of Confidential Information is strictly limited to the preliminary consideration of an investment in the business described in this Document. No portion of this Document may be reproduced, duplicated, or revealed in any manner without prior written consent of the author. Without limiting the generality of the foregoing, the recipient also agrees that photocopying or other duplication of the Document and the data provided without prior written consent is strictly prohibited.

No information contained in this Document or any other written or oral communication transmitted or made available to a recipient of this Document is, or shall be relied upon, as a promise or representation, whether as to the past or future, and no liability will attach except as may be provided in a definitive agreement when, as if it is executed and subject to such limitations as may be provided in such agreement. In addition, any projections and estimates or other forward-looking statements contained in this Document involve numerous and significant subjective assumptions and are subject to risks, contingencies, and uncertainties (many, if not most of which, are outside the control of the author and the Company being represented) which could and likely cause actual results to differ materially. Such assumptions, and risks, contingencies, and uncertainties relate to, among other matters, general business, economic and market conditions.



OUR COMPANY

TAU Express Pte Ltd is an AI Analytics company founded in 2018 as a spin-off from Nanyang Technological University. With deep expertise in machine learning and natural language processing, we built **"TITAN"**, an Intelligence Analytics Platform capable of parsing, extracting and categorizing unstructured text data for intelligent search and analytics.

TITAN sets the foundation for our company's suite of Solutions and has been implemented in projects led by both private and public sectors.



**Industries TAU
has experience in**

- **Banking and Financial Institutions**
- **Aviation**
- **Legal**
- **Healthcare**
- **Logistics**
- **Government**
- **Cybersecurity**

Clients and Partners





Initial Project Lead



Prof LAM Kwok Yan

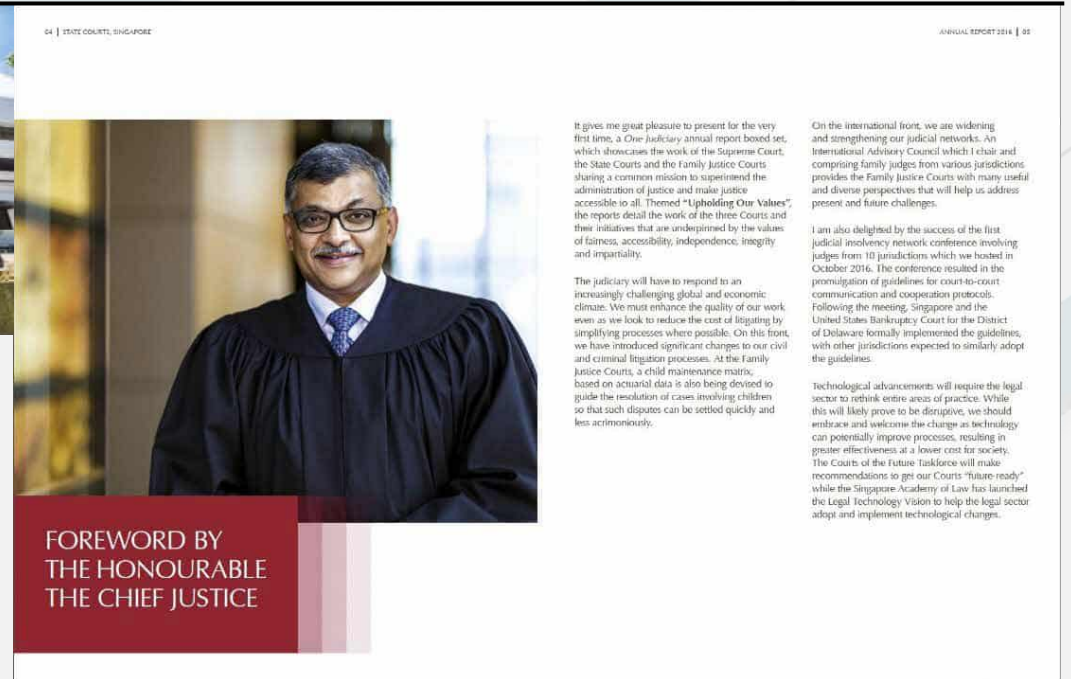
Decades of research excellence, technical expertise and industry experience

Associate Vice President (Strategy & Partnerships) and Professor in School of Computer Science and Engineering at Nanyang Technological University (NTU), Singapore. He is concurrently serving as the Executive Director of the Strategic Centre for Research in Privacy-Preserving Technologies and Systems (SCRIPTS), and Director of NTU's SPIRIT Smart Nation Research Centre.

INTELLIGENT CASE RETRIEVAL SYSTEM FOR DECISION SUPPORT



The Supreme Court and State Courts, together with NTU SPIRIT Centre, have embarked on the development of an Intelligent Case Retrieval System. The aim is to have a more effective retrieval of key legal documents in decided cases for road traffic accidents and personal injury matters. Besides making it easier and quicker for legal research to be performed, it will add to the quality of decision making and judgment writing for judges. The technology developed through this platform may even perform other legal functions, such as Case Summarisation and Logic Reasoning. This will lead the way to a more sophisticated legal research, advisory and decision support system in the future.



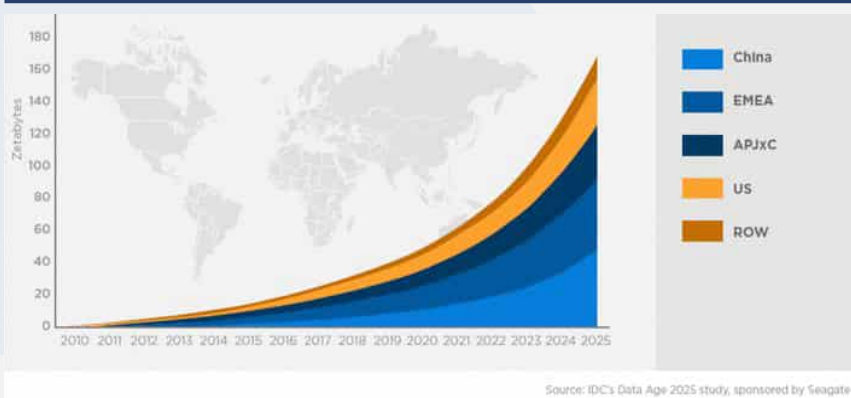
FOREWORD BY THE HONOURABLE THE CHIEF JUSTICE



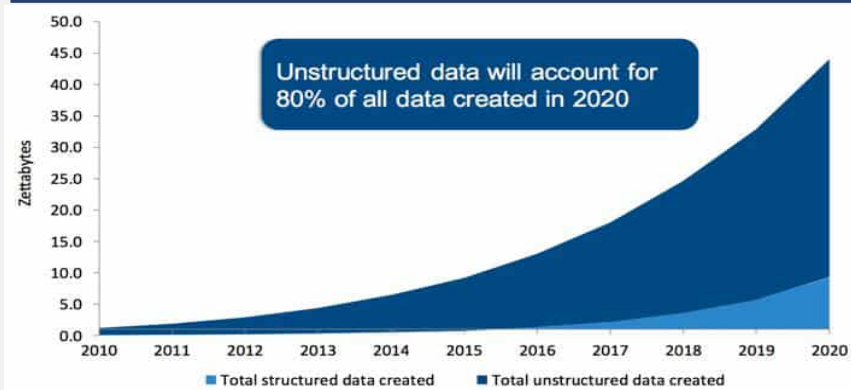


The annual growth of unstructured data has increased by 55-65%

Global Datasphere by Region



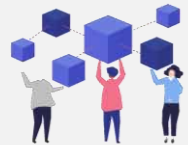
Capacity Growth by Data Type



WHY IS THIS A PROBLEM?



Huge Amount of Unstructured Data - Many key operations depend on personnel handling large amounts of unstructured data in their day-to-day activities and require decisions to be made based on such unstructured data.



Manual Human-dependent Analysis - It is relatively easy to build systems around structured data, but humans are typically required to decipher and make sense of unstructured information.



Opportunity Cost from Unused Data - A lot of unstructured data is stored and rests dormant due to the difficulty of deciphering, understanding, sorting, and searching through the data, which presents a huge opportunity for value extraction.



Inefficient and Prone to Human-error - This results in inefficiencies, lack of consistency in task execution across departments, and human-error.



Key AI Growth Opportunities

AI-POWERED TOOLS TO STREAMLINE TASKS, ENHANCE EFFICIENCY AND DRIVE PRODUCTIVITY GAINS



Automating Data Processing

Process and analyze large volumes of unstructured text quickly and efficiently



Natural Language Processing

Using NLP capabilities for extraction and summarization



Faster Analysis

Rapidly research and generate important insights



Transforming Unstructured Data

Automatically transform unstructured data for further analytics



Knowledge retention

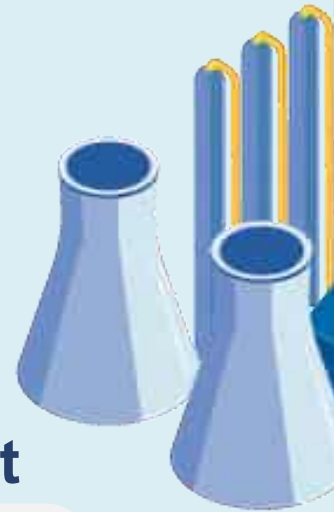
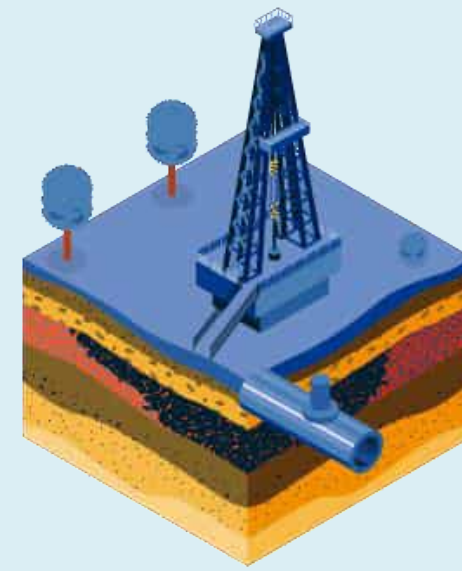
Knowledge representation and engineering to democratize experience



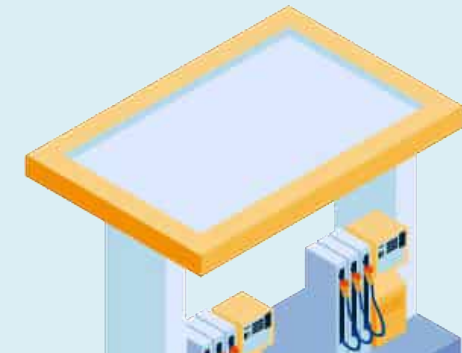
“Data

in the 21st Century is like oil in the 18th Century: an immense, untapped **valuable asset...** those who learn **to extract** and use it will see huge rewards.”

- “Data Is the New Oil of the Digital Economy”, 2020



TITAN is a refinery that turns raw unstructured data into *fuel* for driving business decisions.

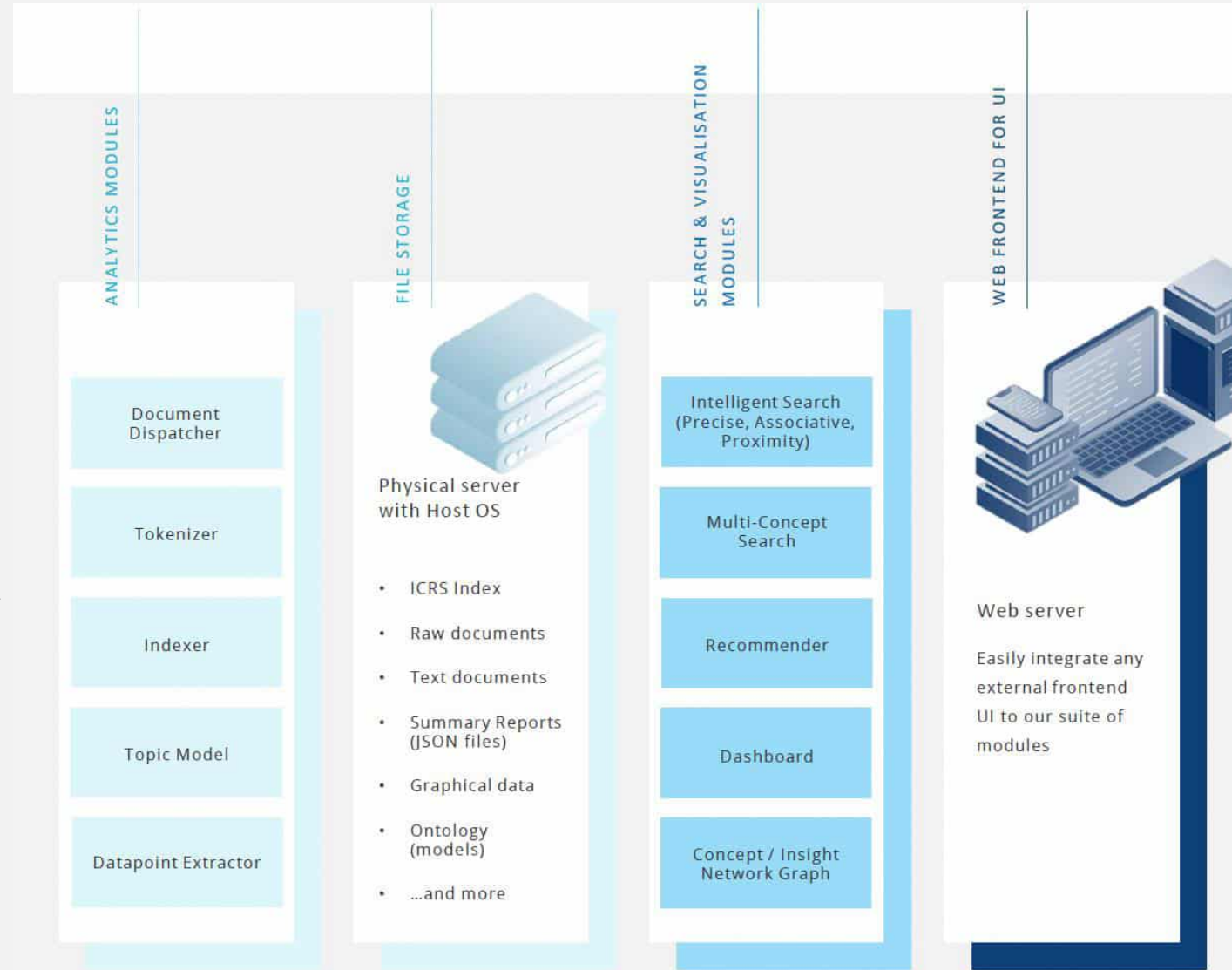




OUR SOLUTION TITAN

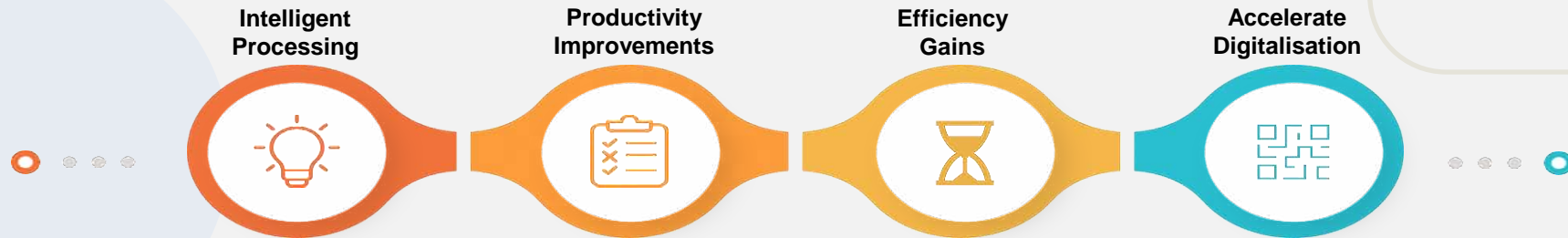
An AI-enabled Platform **proven to achieve >90% accuracy** in retrieving relevant results. An intelligent retrieval system that can parse, extract, and categorize unstructured data to **enable AI-driven data analytics & insights**

- Backed by strong R&D from world class research center
- Industry agnostic
- Modular and scalable architecture
- Designed to support combination of engines/modules to meet evolving requirements





TECHNOLOGY PLATFORM & PRODUCT MODULES



Supporting Smart Nation Initiatives & The Digital Economy

RESTful API for Integration

TITAN

Intelligence Analytics Platform

Intelligent Search

Recommendation Engine

Large Language Model

Query Guidance

Entity Extraction & Population

Automatic Clustering

Automatic Categorization

Fast Asymmetric Topic Modelling

Cluster Visualization



CASE STUDY 1 CYBER THREAT INTELLIGENCE

One of the Top Cyber Security Agency in ASEAN



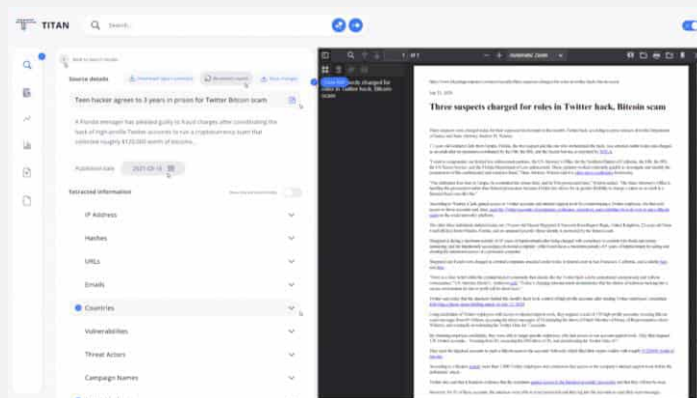
Business Challenges

- **Growing volume & disparate sources** of unstructured CTI data from different vendors
- **Inconsistent terminologies** across different reports
- **Tedious & time-consuming** to manually contextualize cyberattacks
- **Manual evaluation of CTI data** leading to inconsistent and less reliable threat assessments.

Value Creation

- **Increase efficiency** by easily retrieving relevant reports using intelligent search
- **Automated correlation analysis to help** visualizing links and relationships between entities
- **Automated, structured summary** to present users with key indicators/entities at once glance

Intelligent Search



Concept Network Graph



Statistics Dashboard





CASE STUDY 2 LEGALTECH & DOCUMENT INSIGHTS

One of the Top Public Prosecutor Office in ASEAN



Business Challenges

- **Growing number of legal documents** to be reviewed / processed
- **Laborious process** to look through scanned documents to search for required text
- Single document **broken up into multiple files**
- Documents in the form of scanned PDFs are **not searchable or editable**

Value Creation

- **Automatic conversion of scanned PDFs** into textual files to enable intelligent search and document analysis
- **Facilitate efficient drafting** of legal documents
- **Recommend similar clauses** from documents in the repository
- **Reduce time and cost** in looking for relevant documents

Intelligent Search

Query Expansion

Original Words	Associated Words
most favoured nation	most favoured nation status, most favoured nation, trade agreement

Search Results

File Name	Document Title	Snippet	Date of Signature
EFTA - Hong Kong, China/tpo-18	APPENDIX 2 TO ANNEX X LIST OF RESERVATIONS OF ISLAND	4 (Most Favoured Nation Treatment), Article 3	2011-08-21
EFTA - Hong Kong, China/tpo-25	APPENDIX 4 TO ANNEX X LIST OF RESERVATIONS OF NORWAY	4 (Most Favoured Nation Treatment), Article 3	2011-08-21
EFTA -	Agreement on Agriculture between Canada and the	applied most favoured nation tariff in effect at that time	2008-01-26

Document/Clause Comparison

Clause Recommender

Clause Recommender

Search the repository for clauses used in similar context documents

establishing a free trade area without exclusions and restrictions, which in mutual trade provides for the recourse of anti-dumping, compensating and special protective measures on the basis of pursuing a uniform policy in tariff and import regulation, uniform rules of competition, application of subsidies and other forms of state support;

AGREEMENT TITLE	RECOMMENDED CLAUSE	MARK RELEVANCE	MARK AS MOST RELEVANT
Common Economic Zone (CEZ)000-1	• establishing a free trade area without exclusions and restrictions, which in mutual trade provides for the recourse of anti-dumping, compensating and special protective measures on the basis of pursuing a uniform policy in tariff and <u>import</u> regulation, uniform rules of competition, application of subsidies and other forms of state support;	Relevant <input type="checkbox"/>	<input type="radio"/>
Common Economic Zone (CEZ)000-1	• executing a set of measures to complete the development of a free trade area without exclusions and restrictions that fosters the free movement of goods and services under uniform rules of competition and application of subsidies and the pursuance of a concerted policy of tariff and non-tariff regulation;	Relevant <input type="checkbox"/>	<input type="radio"/>




Retrieval Augmented Generation

The screenshot shows a user interface for Retrieval Augmented Generation. It is divided into two main sections: 'Dataset' and 'Output'.
1. **Dataset:** This section contains two radio buttons: 'IMDA Regulations & Guidelines' (selected) and 'Open-Source CTI Reports'.
2. **Input Query:** A text input field contains the query: 'what does the PDPA govern and what is the penalty for breach of user privacy under the act?'. Below the input is a checked checkbox labeled 'Detailed Reply'. A 'Submit' button is located at the bottom of this section.
3. **Output:** This section displays the generated response. The text reads: 'The Personal Data Protection Act (PDPA) governs the collection, use, disclosure, and protection of personal data in Singapore. The Act applies to all organizations that collect, use, or disclose personal data, including government agencies and private companies. Under the PDPA, there are specific provisions for the handling of sensitive personal data, such as information about an individual's race, religion, health, or sexual orientation. Organizations must obtain explicit consent from individuals before collecting, using, or disclosing their sensitive personal data, and they must also ensure that such data is protected with appropriate security measures. The penalty for breach of user privacy under the PDPA can be severe. If an organization fails to comply with the Act's provisions and causes harm to individuals as a result, it may be liable for damages and face fines of up to \$1 million. In addition, the Commissioner of Personal Data Protection has the power to impose penalties of up to \$50,000 on individuals who breach the Act's provisions. Repeat offenders may also be subject to criminal prosecution and imprisonment.'
4. **References:** This section lists four references, all pointing to 'raw_pdfs\Personal Data Protection Act 2012.pdf' at different page numbers: 89, 32, 58, and 111.

- 1 Selected documents stored in repository for access and retrieval of relevant knowledge during model's inferencing phase.
- 2 User-entered prompt. A more detailed response will be generated if "Detailed Reply" is checked.
- 3 Model learns to extract relevant information from the documents and uses it to generate answers.
- 4 Knowledge base in the repository can be updated regularly to ensure that the language model remains up-to-date with the latest information.



The context retrieval ensures that the model generates responses that are grounded in information contained within the knowledge base.

 TAU EXPRESS

Dataset:

IMDA Regulations & Guidelines Open-Source CTI Reports

Statutes **Pharmaceutical Drugs**

Input Query

Detailed Reply

Low Similarity Threshold

Submit

Output

References



THANK YOU

We look forward to working with you



77 Science Park Drive #04-14
Singapore, 138632



info@tauexpress.com



<https://tauexpress.com>

